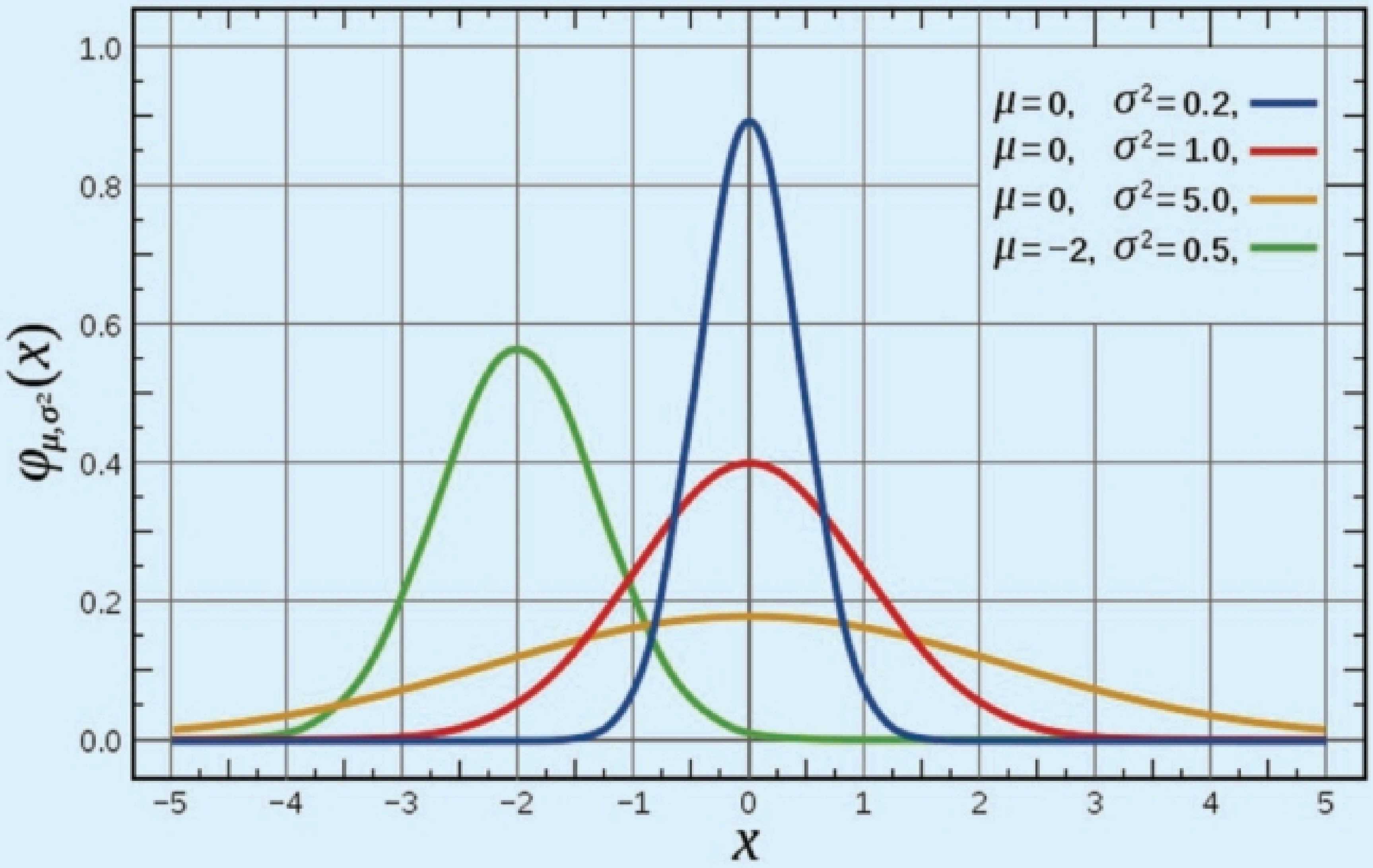
مقدمة في
الإحصاء الزراعي



تأليف
بوب ديفز

ترجمة
أ.د. مهدي بن معيض السلطان

جامعة الملك سعود
النشر العلمي و المطابع

بسم الله الرحمن الرحيم

مقدمة في الإحصاء الزراعي

تأليــف

بوب ديفز

ترجمة

أ.د. مهدي بن معيض السلطان

النشر العلمي والمطابع - جامعة الملك سعود

ص.ب ٦٨٩٥٣ - الرياض ١١٥٣٧ - المملكة العربية السعودية

هذه ترجمة عربية مصرح بها من مركز الترجمة بالجامعة لكتاب :

Introduction to Agricultural Statistics
By: Bob Davis

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

ديفز، بوب

مقدمة في الإحصاء الزراعي ؛ بوب ديفز ؛ مهدي بن معيض السلطان.

- الرياض، ١٤٣١هـ

٤٢٠ ص، ١٧ × ٢٤سم

ردمك : ٦ - ٩٠٥ - ٥٥ - ٩٩٦٠ -٩٧٨

١- الزراعة - احصائيات ٢- الإنتاج الزراعي - احصائيات

أ. السلطان، مهدي بن معيض (مترجم) ب. العنوان

ديوي ٣٣٨,١٠٥ ١٤٣٢/٩٠٦٦

رقم الإيداع ١٤٣٢/٩٠٦٦

ردمك : ٦ - ٩٠٥ - ٥٥ - ٩٩٦٠ -٩٧٨

حكمت هذا الكتاب لجنة متخصصة، شكلها المجلس العلمي بالجامعة، وقد وافق المجلس على نشره - بعد إطلاعه على تقارير المحكمين- في اجتماعه الثاني للعام الدراسي ١٤٣٢/١٤٣١هـ المعقود في تاريخ ١٤٣١/١٠/٢٤هـ الموافق ٢٠١٠/١٠/٣م.

النشر العلمي والمطابع ١٤٣٢هـ

إهـداء المترجم

إلى كل مسلم حمل همّ العلم، وكان تقدّم أمّة محمد صلى الله عليه وسلم هدفه.

أهدي هذا الجهد المتواضع،،،،،

المترجـــــم

مقدمة المترجم

الحمد لله رب العالمين والصلاة والسلام على أشرف الأنبياء والمرسلين سيدنا ونبينا محمد وعلى آله وصحبه وسلم ... وبعد :

الإحصاء كعلم له تطبيقات كثيرة في شتى مجالات المعرفة ، ويعتمد هذا العلم بصفة أساسية على الرياضيات وبذلك يمكن عرض الموضوعات بمستويات مختلفة تتلاءم مع نوعية القارئ ودرجة معرفته بالرياضيات.

ويعتبر القطاع الزراعي أحد مجالات العلوم التطبيقية لعلم الإحصاء الاّ أن المراجع العربية في هذا المجال قليلة مما يعيق استعانة الطلاب والمبتدئين بهذا العلم في الدراسة والبحث والتحليل. ونتيجة لذلك تم البحث عن كتب باللغات الأخرى للاستفادة من تجارب الآخرين وقد يسر الله لنا العثور على هذا الكتاب وتم عرضه على مركز الترجمة بجامعة الملك سعود وتمت الموافقة على ترجمته ليكون مرجعاً علمياً في هذا المجال.

وها أنا أضع بين يدي القارئ العربي هذا الجهد الذي لا يخلو من التقصير والذي احتوى على ثلاثة عشر فصلاً شاملة لتطبيقات إحصائية متعددة في المجال الزراعي.

ولا يفوتني أن أتقدم بخالص الشكر والعرفان لكل من ساهم معي في إخراج هذا الكتاب إلى حيّز النور وأخص بالذكر جامعة الملك سعود ممثلة في مركز الترجمة

وأسرتي التي ضحت بالكثير من الوقت والجهد في سبيل إنجاز هذا العمل الذي كان على حساب الكثير من متطلباتهم .

وإنني إذْ أتقدم بهذا الجهد المتواضع للقارئ العربي أرجو من الله سبحانه وتعالى أن يتقبله مني عملاً صالحاً فهو من وراء القصد وهو الهادي إلى سواء السبيل.

المترجـــــم

مقدمة المؤلف

صمم هذا الكتاب للتدريس في فصل دراسي واحد بوصفه مقدمةً للطرق الإحصائية المعتادة التي تدرّس لطلبة العلوم الزراعية في مرحلة البكالوريوس باستخدام أمثلة وتطبيقات ذات علاقة يمكن فهمها بسهولة من قبل الطالب. وقد تم عرض محتويات الكتاب بتسلسل منطقي مما يساعد المحاضر بالمضي من أول فصل إلى آخر فصل عند توفر الوقت اللازم.

من جهة أخرى فعند ضيق الوقت يمكن حذف الفصول الأخيرة دون التأثير على معظم المفاهيم المهمة. استخدم الشرح أمثلة زراعية لتعليم الأفكار والمفاهيم الإحصائية. وتجدر الإشارة إلى أن صياغة النظرية الإحصائية تمت بشكل قصصي في حين تم عرض المفاهيم باستخدام رياضيات مبسطة. المشكلات ذات الطبيعة الزراعية المعروضة بنهاية الفصل يمكن استخدامها لتعزيز الأفكار بالنسبة للطالب ، وقد اشتملت الفصول على عرض التحليل الإحصائي باستخدام برنامج الجداول الإلكترونية على الحاسب الشخصي عندما يكون ذلك عملياً باستخدام بعض الأمثلة.

محتويات الكتاب

اشتمل الفصل الأول على مقدمة عن الإحصاء والاستدلال الإحصائي مع ملحق عن المجاميع في حين غطى الفصل الثاني المقاييس المعتادة لقياس النزعة المركزية

والتشتت المستخدمة في تلخيص البيانات المبوبة وغير المبوبة. بعد هذه المقدمة تم التطرق للاحتمالات والتوزيعات الاحتمالية حيث تم عرض مفاهيم الاحتمالات الأساسية اللازمة لدراسة التوزيعات الاحتمالية. وطرق حساب الاحتمالات للتوزيعات الأكثر استخداماً مع التركيز على استخدام الجداول الملحقة. بعد ذلك تم تقديم المعاينة وتوزيعات المعاينة، حيث إن إتقان الطالب لفكرة توزيع المعاينة للمتوسط على سبيل المثال يسهّل عليه فهم بقية موضوعات الاستدلال الإحصائي. من ناحية ثانية فإن الفصول من الثالث حتى السادس تشتمل على المفاهيم الأساسية للمقرر، وذلك بعرض الأسس اللازمة للفصول اللاحقة.

وقد بدأنا موضوعات الاستدلال الإحصائي بفترات الثقة في حالة العينة الواحدة ثم توسيع ذلك لحالة عينتين. تم بعد ذلك عرض اختبارات الفروض لنفس الحالات الواردة في فترات الثقة مما يمكّن الطالب من ملاحظة العلاقة بينهما. ثم اشتملت الموضوعات على نماذج تحليل التباين باتجاه واحد واتجاهين والتي يساهم الحاسب الآلي في تسهيل فهم هذه الموضوعات. مع التركيز على تفسير النتائج. اختبارات الفروض للنسب لثلاث عينات أو أكثر تم عرضه باستخدام مربع كاي، إضافة لاختبار الاستقلال الإحصائي وتمثيل التوزيعات باستخدام مربع كاي.

في الفصل الحادي عشر تم عرض توفيق المنحنيات حيث تم شرح الارتباط البسيط والانحدار الخطي البسيط بالتفصيل. وفي ذلك الفصل أيضا تم شرح الانحدار المتعدد ثم السلاسل الزمنية تلا ذلك عرض بعض الموضوعات المهمة في إدارة الأعمال الزراعية و توفيق الاتجاه الزمني الخطي وغير الخطي باستخدام نموذج الانحدار ثم ختم الفصل بالتغيرات الموسمية في السلاسل الزمنية.

اشتمل الفصلان الأخيران على طرق الاختبارات اللامعلمية والتي من أهمها اختبار مان وتني واختبار الإشارة واختبار التتابع وارتباط الرتب، وكذلك الأرقام

القياسية وبعض الموضوعات المهمة لطلاب إدارة الأعمال الزراعية.

وبعد دراسة مقرر الإحصاء اعتماداً على هذا الكتاب فإنه يتوقع أن لا يقتصر فهم الطالب على الطرق الإحصائية ولكن نأمل أن يتعدى الأمر ذلك بحيث يستطيع الطالب تحليل معظم المشكلات بالقطاع الزراعي تحليلاً إحصائياً.

شكر وتقدير

أتقدم بالشكر لزوجتي أنيتا على تشجيعها وتأييدها لي خلال مشروع تأليف هذا الكتاب ودراستي الجامعية في SWT ، وكذلك أقدم شكري لمن كان لديهم الرغبة في مناقشة كيفية استخدامنا للإحصاء في مجالات عملهم وكذلك لمن أمدنا ببعض المعلومات والبيانات التي ساعدتنا في صياغة المسائل الواردة في هذا الكتاب.

المـؤلـف

المحتويات

الصفحة

الفصل الثاني: تلخيص البيانات

الفصل الثالث: الاحتمال

الفصل الرابع: التوزيعات الاحتمالية

الفصل الخامس: المعاينة وتوزيعات المعاينة

الفصل السابع: التقدير الإحصائي لعينتين

الفصل الثامن: اختبار الفروض

الفصل التاسع: تحليل التباين

الفصل العاشر: تطبيقات مربع كاي

الفصل الحادي عشر: الارتباط، الانحدار، والسلاسل الزمنية

الفصل الثاني عشر: الإحصاءات اللامعلمية

الفصل الثالث عشر: الأرقام القياسية

الفصل الأول

مقدمـــــة

Introduction

يهدف هذا الكتاب إلى إعطاء مقدمه في الاستدلال الإحصائي في مجال الزراعة وعلـوم الحياة. وعلى الرغم من أن قارئ هذا الكتاب قد لا يكون إحصائياً إلاّ أنه عنـد الانتهـاء منه سيكون لديه القدرة على فهم مادة الموضوع ومن ثم استخدام معظم طرق التحليل الإحصائي المعتادة مع قدرته على مناقشة المسائل المعقدة بمهارة.

ما هو الإحصاء؟ ربما يكون المفهوم الأكثر شيوعاً للإحصاء بأنه يعبر عن بيانـات عدديـة استـخدمت كـسجلات للماضـي وبـذلك فـإن الإحـصاء كبيانـات عدديـة يمكـن استخدامه للتعبير عن عدة مفاهيم سواء في الوقت الحاضر أم المستقبلي بناءً على طبيعـة تلـك المفـاهيم ومـدلولاتها بالنـسبة للـشخص المهـتم بهـا. وكمثـال علـى ذلـك ، خـلال مواسم لعب كرة القدم ما هي المعلومات التي تكتب في الصحف اليومية عـن المباريـات تحت العمود المسمى إحصاءات وما مدى أهمية تلك الإحـصاءات؟ تعتمـد أهميـة تلـك البيانات على من يقرأها فإن كان القارئ لها هو مدرب الفريق المنـافس فـسيهتم بأعـداد التمريرات والأخطاء ، ...إلخ في حين أن القارئ العادي مهتم بمعرفة نتيجة المباراة.

أمّا في حالة البيانات الزراعية فإن تاجر تربية المواشي يهـتم بمعرفـة أعـداد القطيـع في المرعى والأعداد التي تم إحلالها الأسبوع الماضي ، متوسط سعر علائق التغذيـة.. إلخ وبذلك يستطيع بناءً على ذلك إدارة مشروعه.

وفي هذا السياق فإن العوامل الأساسية التي تحكم ملاءمة الإحصاءات من عدمها تعتمد على معرفة الشخص بالبيانات ومدى مناسبتها بالنسبة للقرارات المستقبلية. ويعرف الإحصاء الذي يهتم بالأرقام والتعامل معها بالإحصاء الوصفي Descriptive Statistics أما الاستدلال الإحصائيStatistical inference فيشير إلى عمليات صنع القرارات حيث يتضمن استخدام المنطق وكذلك الطرق العلمية في صناعة القرارات، حيث يركز على حل المشكلات وكما نعلم فإن المنطق يتضمن الأسباب من الخصوصية إلى العموم. في حالة الاستدلال الإحصائي فإن المعتاد هو جمع المشاهدات من وحدات صغيرة (عينة) من المجتمع وتحليل تلك البيانات باستخدام طرق التحليل المعتادة ثم الاستدلال بها نظراً لأن النتائج المتحصل عليها من العينة يجب أن تكون هي نفسها المتحصل عليها من جميع الوحدات (المجتمع).

وكمثال على ذلك فإن البث الإعلامي في الولايات المتحدة الأمريكية يتنبأ بمن سيربح الانتخابات العامة للرئاسة قبل إغلاق عملية الاقتراع في الشاطئ الغربي. وتتم هذه العملية بسؤال عينه مختارة بعناية من الناخبين كيف تتم عملية تصويتهم ثم يتم الاستدلال من النتائج المتحصل عليها؛ نظراً لأن ذلك التصويت سوف يتم بنفس الطريقة من قبل المجتمع ككل وهذا ما يسمى بالمنطق الاستقرائي.

الطريقة العلمية تصف الإجراء المتبع في العلم لعمل التجربة. والذي يشتمل عدة خطوات تتمثل في تعريف المشكلة، وصياغة الفروض، وتحديد البيانات المطلوب جمعها وتحليلها، واختيار طريقة التحليل المناسبة ثم تحديد ملاءمة نتائج التحليل للمشكلة المدروسة. و يمكن التساؤل كيف يمكن للاستدلال الإحصائي أن يساهم في هذا الإجراء ؟

بعد صياغة المشكلة والفروض المناسبة يمكن اتخاذ القرار المناسب للاختبار- ما

هي البيانات اللازمة ، كيف يمكن الحصول عليها وتنظيمها وتحليلها بالطريقة التي تعطي إجابة مناسبة للمشكلة محل الدراسة - ويعتبر الاستدلال الإحصائي إحدى الطرق الإحصائية للقيام بهذه المهمة ولكن يتم اختيار الأدوات اللازمة طبقاً لطبيعة المشكلة وطريقة التحليل الملائمة لها ونوع الخلاصة التي نرغب في الوصول إليها. وبصفة عامة إذا أردنا التعبير عن بعض الخلاصات في صورة صيغ احتمالية أو عندما نحاول التعبير عن درجة التأكد من شيء ما بحس احتمالي ، فإنه يجب تطبيق بعض صيغ الاستدلال الإحصائي.

ربما الآن نستطيع أن نعرّف الإحصاء ، يشتمل الإحصاء على تطوير وتطبيق الطرق والأساليب لجمع وتحليل وتفسير البيانات العددية ، لذا فإن الاعتماد على النتائج المستخلصة من هذه العملية يمكن تقييمها بصورة موضوعية باستخدام التعابير الاحتمالية. والجزء الأخير من التعريف– استعمال الاحتمالات لتقييم موضوعية النتائج والاستدلال بناء على البيانات – هي المساهمة الوحيدة والرئيسة للإحصاء في الاستدلال الاستقرائي.

الاستعمالات الخاطئة الشائعة للإحصاء Common Misuses of Statistics

يقول ديسريلي Disreali بأن هناك الكذب ، الكذب الملعون ، والإحصائيين مما يجعل الإحصائيين مع بعضهم في الشركة غير مرغوب فيه. وقد كتب هوف Huff كتاب صغير مرتب أسماه ، *كيف تكذب باستخدام الإحصاء* ، كما فسر ريتشمان Reichmann الاستعمال السيئ للإحصاء. والحقيقة القائلة بأن الآخرين يصلون إلى نفس الملاحظات تقودنا للاعتقاد بأنه ليس كل من يستخدم التحليل الإحصائي سوف يقوم بذلك بنفس الدرجة من الاهتمام أو الانتباه. ففي التعاملات التجارية وكذلك

الحال في حياتنا الشخصية يتم الدفاع باستخدام العروض المدعمة بالإحصاءات. لذا فإن أحد الأسباب لتعلم الإحصاء هو للحماية الذاتية! في الجزء التالي سوف نتطرق لمعظم الطرق الشائعة التي تؤدي للتضليل عند استخدام الإحصاء.

التحيز Bias

في الغالب يكون من السهل اكتشاف التحيز الواضح أو المحاباة في الإعلانات التي تستشهد بالإحصاءات لإثبات تفوق منتج معين، في حين أن إعلانات المنافسين تستشهد بإحصاءات أخرى لإثبات العكس. ولكن التحيز غير المقصود ينسلّ للعمل الإحصائي نتيجة التحيز في البيانات المجموعة باستخدام الاستبيانات أو التصويت الناتج من النقص أو الخطأ في طريقة المعاينة المستخدمة، والضعف في تصميم الاستبانة، والأسئلة الايحائية، أو من الباحث الذي يحاول/تحاول إثبات فكرته/فكرتها. وفي حالة استخدام البيانات في أي دراسة فإن المتوقع أن يشير الباحث لمصدر البيانات وعدم ذكر ذلك يوحي بأنه ربما يكون هناك مشكلة.

التعميمات الخاطئة Faulty Generalizations

نرى أحياناً التعميمات الخاطئة نظراً لكون العينة التي تم اختيارها كانت صغيرة جداً، مما يجعل من غير الممكن تمثيل المجتمع على نحو كافٍ بناءً على النتائج التي تم استخلاصها. فمثلاً، من الصعوبة تمثيل جميع أسماك فصيلة معينة بصورة عادلة في البحيرات العظمى باستخدام عينة مكونة من خمس سمكات. والأكثر سوءاً التعميم باستخدام عينة حجمها واحد والتي تستخدم في الغالب. والتعميمات الخاطئة الأخرى تأتي من العينات غير الممثلة مثل، العينات التي تكون كافية في حجمها ولكنها تختلف

عن المجتمع في بعض التقديرات المهمة ، فمثلاً دراسة استقصائية عن خريجي الكلية من دفعة ١٩٤٠م أشارت بأن متوسط رواتبهم كانت ٩٥١٠٧ دولارات سنوياً ، وهذه الدراسة مبنية على عدد ٢١ استجابة من عينة الدراسة البالغ عددهم ١٤٢. وبذلك تكون هذه النتيجة خاطئة إذا كان أولئك الذين استجابوا للدراسة لهم دخول عالية أكثر من الآخرين أو ربما أنهم قاموا بتجميع كل مكتسباتهم كما يحدث في مثل هذه الحالات. أيضاً إذا كان بعض الخريجين الذين استجابوا للدراسة يتقاضون دخول عالية جداً فإن ذلك سوف يؤدي لتضخيم المتوسط بدرجة كبيرة جداً.

الاستدلال الخاطئ Faulty Deduction

الخطأ الشائع للاستدلال الخاطئ يحدث من القيام بتطبيق قاعدة عامة صالحة كما لو كانت قانون ثابت. لذا فالتقرير الحكومي الختامي المتضمن بأن سعة مراعي الغزلان أكبر من الطلب بـ ١٠,٢٪ ربما تكون صحيحة على مستوى الدولة ولكن ليس من الضروري أن يكون ذلك صحيح لكل منطقة. فقد يكون هناك عجز أو نقص في الغرب وسعة زائدة بنسبة ١٥,٥٪ في الجنوب الشرقي. ولكن إذا أمكن صياغة التقرير كاتجاه عام يسمح بالاختلافات الفردية فإن هذه المشكلة قد لا تكون مزعجة.

البيانات غير المتماثلة Noncomparable data

المقارنات تتم عادة بين وحدات غير متشابهة فعلاً. فمثلاً أشار ممثل نقابة السيارات بأن سعر سيارة العائلة السيدان ذات الأربعة أبواب ارتفع تقريباً ٥٠٪ في آخر عشر سنوات. وقد أهمل الإشارة إلى أنه خلال تلك الفترة قام المصنع بإضافة العديد من الخصائص للسيارات مثل التجهيزات القياسية التي تجعل السيارات الجديدة تختلف

عن القديمة. والخطأ الناتج عن عدم التماثل يؤثر في مؤشر السعر بشكل عام ؛ نظراً لأن التقنية تتغير باستمرار والمواصفات الخاصة تتغير من فترة زمنية لأخرى.

الأخطاء في الدلالات Errors in Semantics

تستخدم أحيانا الكلمات الشاذة أو الملمّعة للتأثير على المستمعين كما في الحملات السياسية. لذا فإن الاحتراس واجب من الأسئلة في الاستبيانات التي ربما تقود أو توحي بإجابة معينة. فمثلاً ، سؤال " هل تحب الشرطة" سوف يعطي إجابة تختلف قليلاً فيما لو تمت صياغة السؤال على النحو " ما هو رأيك في قسم الشرطة".

افتراض السببية نتيجة الارتباط Assuming causation from correlation

تعني هذه العبارة أنه نظراً لأن بعض الأشياء تسبق الآخر في الحدوث عبر الزمن فقد نفترض أن الأول يسبب الثاني. فمثلاً ، افترض أنك تعرضت للعواصف الرعدية ومن ثم أصبت بنزلة برد. فإنك تفترض أن تعرضك للمطر هو سبب إصابتك بالبرد وليس بالضرورة أن يكون ذلك. بصفة عامة إذا كان العاملان A وB تتذبذب معاً (A وB مرتبطين) فربما أن A تسبب B ولكن أيضا قد يكون B تسبب A ، أو أن A و B تؤثر في بعضها بصفة مستمرة ، أو بشكل متقطع ، أو ربما A و B تتأثر بعامل آخر C أو ربما يكون الارتباط نتيجة الصدفة. وعليه فقد تكون تعرضت لفيروس برد غير معروف والذي هاجمك أثناء تعرضك للمطر. ولا يمكن استخدام الإحصاء للاستدلال عن السببية الناتج عن الارتباط. وإذا تم عمل مثل ذلك الاستدلال فإنه يجب أن يكون مبني على حكم مهني للباحث أو الباحثة والذي يعرف عن العوامل من واقع مهنته. وكل ما يقوله الإحصاء في هذه الحالة إن العاملين يتحركا معاً بطريقة معينة.

التبسيط الزائد Oversimplification

تحـدث الأخطـاء المعتـادة نتيجـة التبـسيط الزائـد للموضـوع وإهمـال الـشروط الـضرورية. وقـد تكـون الحقيقـة المعروضـة صـحيحة ولكـن ربـما تـضلل الجمهـور إذا تم إهمـال بعـض الحقـائق ذات الـصلة بهـا. فمـثلاً ، الإعـلان للجـرار الحـالي يوضـح بـأن مبيعاته زادت بنسبة ١٥٥٪ في ولاية تكساس. ولكن من متى؟ ولذلك فإن المقارنة ليس لها معنى لعدم إعطاء معلومات عـن سـنة الأسـاس. وعمومـاً فـإن أصـحاب الـدعايات والسياسيين وآخرين يستخدمون هذا التكتيك أو هذه الطريقة باستمرار.

الدقة الزائفة Spurious Accuracy

نحن نعيش الآن في عـصر الكمبيـوتر حيـث يمكـن إنـشاء إجابـات بعـدد كـبير مـن الفواصل العشرية. ولكن لسوء الحظ فإن الإجابة لن تكون أكثر دقة من البيانات. فلو تم تقريب البيانات لأقرب ١٠٠ دولار، كما في حالة بيانات الدخل، ففي هـذه الحالـة لـن يتم كتابة متوسط الدخل للدراسـة علـى أنـه ٤٤٢٠٣,٥٠ دولاراً. وقـد يبـدو ذلـك مثـير للإعجاب ولكن تلك دقة زائفة. وحيث إن البيانات في هذه الحالة تم تقريبها لآخر ١٠٠ دولاراً فإنه يجـب كتابتهـا علـى النحـو ٤٤٢٠٠ دولار. وقـد يـتم تـشويه المعنـى الحقيقـي للحقائق ببساطة. لـذا يجـب علـى الباحـث الإحـصائي أن يحـاول تجنـب العـرض المـشوّه للحقيقة ويكتشف الخطأ في استخدام الإحـصاء مـن قبـل الآخـرين وبـذلك فـإن موقـف النقد يعتبر أساسياً.

نظرة للأمام A Look Ahead

الموضوعات التي ستتم تغطيتها في بقية الكتاب مقسمة لثلاثة أجزاء هي:

١- استخدام الإحصاء في تلخيص البيانات ممّا يمكّن من فهمها ببساطة.

٢- الاحتمالات، والتوزيعات الاحتمالية، والمعاينة، وتوزيعات المعاينة والتي تتضمن الخلفية المطلوبة للموضوعات في الجزء الأخير.

٣- الاستدلال الإحصائي.

نحن نصادف الإحصاء الوصفي يومياً في تعاملاتنا الروتينية. فمثلاً شخص معين يعطينا دائماً المتوسط لشيء معين وآخر قد يخبرنا بمتوسط إنتاجية المحاصيل أو متوسط المحتوى الرطوبي لآخر حمولة للحبوب أخذت من المخزن...إلخ. ولذلك نستطيع ببساطة فهم ماذا تعني عندما يقوم شخص ما بعرضها لنا وكذلك عندما نرغب في استخدامها كرقم أو رقمين لوصف البيانات. الاحتمالات والتوزيعات الاحتمالية...إلخ تستخدم كمعلومات أساسية لفهم الاستدلال الإحصائي؛ لأن معظم التحليل الإحصائي تقريباً مبني على الاحتمال. وسوف يساعد قاموس المصطلحات في آخر الكتاب في فهم المصطلحات غير المألوفة خلال دراسة الموضوعات. ويتركز الاهتمام في هذا الجزء على المفاهيم الأساسية والتي سيتم استخدامها في الجزء الأخير في التقدير واختبارات الفروض.

الموضوعات في الاستدلال الإحصائي توضح طرق تحليل البيانات ولذلك نستطيع تحديد ما إذا كانت إحدى المعالجات تختلف فعلاً عن الأخرى أم لا. ويمكن تحديد ذلك باستخدام تعبير احتمالي حول ذلك الاختلاف إذا كان الاختبار الإحصائي يشير لوجود اختلاف بينهما. بينما في بعض الحالات يمكن النظر للبيانات ونجد أنها نفسها و لا يوجد اختلاف، وفي حالات أخرى فإن ذلك ليس مهماً. وقد تم تطوير الطرق الإحصائية خلال نصف القرن الماضي والتي تتخذ من التخمين محاولة لوجود الاختلاف من عدمه. وسيتم قضاء بعض الوقت لتعلم هذه الطرق. وهناك بعض الطرق المعتادة كل منها مخصص لإجراء اختبار معين وسوف يتم التطرق لها جميعاً.

المراجع References

Huff, D. *How to lie with statistics*. New York: W.W. Norton, 1954.
Ltd., 1961. Reichmann, W.J. Use and absue of Statistics. London: Methuen

الملحق: المتغيرات ومتغيرات التجميع
Appendix: variables and Summations variables

المتغيرات variables

يعرف المتغير كأي شيء والذي يمكن أن يكون له مدى من القيم. في الإحصاء يتم التعبير عن المتغيرات بالرموز المعتادة للحروف الهجائية مثل X و Y . وهذا من الأشياء الروتينية المتعارف عليها. ويمكن ببساطة إعطائها أسماء أو استخدام الحروف الإغريقية ولكن ذلك يتطلب استخدام نوع معين من الدلالة. ونظراً لأن المتغيرات تأخذ مدى من القيم فإنه يمكن تعريفها بناءً على أنواع القيم المفترضة إلى نوعين.

المتغيرات الوصفية: هي المتغيرات التي لا يمكن قياسها حيث تعبر عن بعض الصفات التي لا يمكن قياسها عددياً مثل الحالة الزوجية، الأجزاء الجيدة أو المعيبة، أنواع الفواكه، فئات العمر ...إلخ. ويتم التعبير عن هذا النوع من المتغيرات باستخدام قيم افتراضية لها مثلاً تكون X=1 إذا كانت الكرة سوداء وصفر عدا ذلك، أو Y=1 للأنثى و ٢ للذكر ...إلخ.

المتغيرات الكمية: تعبر عن الأشياء التي يمكن قياسها أو عدّها. فإذا كان المتغير يمكن عده فقط فإنه يكون منفصل أو متقطع.

المتغيرات المنفصلة: تكون قيمها أعداد صحيحة. وهناك العديد من مثل هذه المتغيرات في الإحصاء مثل أعداد العجول الذكور في حظيرة سعتها ثمانية، عدد مرات الوصول لمنطقة العرض خلال فترة زمنية معينة، أو عدد مرات النجاح في عدد n محاولة لتجربة ما.

المتغيرات المتصلة: هي التي يمكن قياسها ولذلك يمكن أن تأخذ أي قيمة في المدى للمتغير بناء على دقة أدوات القياس المستخدمة. فمثلاً إذا كانت X تعبر عن إنتاج السماد بالطن في اليوم من أحد مصانع الأسمدة فإن قيمة X يمكن أن تتراوح بين الصفر والسعة القصوى للمصنع والتي نفترض أنها تساوي ١٠٠ طن. وإنتاج الأسمدة متغير مستمر ولذلك يمكن استخدام وحدات القياس التي نرغبها: ١٠٠طن، أطنان، عشرات الأطنان، مئات الأطنان، آلاف الأطنان، ...إلخ، وخاصة إذا كنا نستخدم الميزان الإلكتروني.

التجميع Summations

تستخدم الصيغ في الإحصاء كطرق عامة للتعبير عن كيفية حساب المتغيرات محل الاهتمام. فمثلاً ، الصيغة المستخدمة لحساب المتوسط الحسابي لمجموعة من البيانات توضح كيفية الحساب بغض النظر عن حجم البيانات التي تم اختيارها. ولذلك فإنه لا يجب تغيير تلك الصياغة عند تغير حجم أو نوع البيانات موضع الدراسة. وقد تحتوى الصيغ على علامة التجميع والتي تساعد في تحديد كيفية إجراء الحسابات.لذا سيتم عرض جبر المجاميع في الجزء التالي.

مؤشر التجميع Index of Summation

رمز التجميع Σ يعني إضافة القيم للمتغير بدءاً بأول قيمة في المؤشر مروراً بكل القيم ثم الانتهاء بالقيمة الأخيرة في المؤشر.وتكتب أول قيمة في المؤشر أسفل الرمز Σ والقيمة الأخيرة تكتب أعلى الرمز Σ . فمثلاً يتم تقدير المصطلح في المعادلة رقم (1.1) بإضافة القيم العشر لـ X_i .

$$\sum_{1}^{10} X_i \qquad (1.1)$$

ابتداء بـ X_1 ثـم الاسـتمرار بـالقيم X_2 ثـم X_3 وهكـذا... حتـى ننتهـي بالقيمـة X_{10}. وحيث إن لـدينا عـشر قـيم في مؤشـر التجميـع قمنـا بتجميـع القـيم العـشر الأولى لـ X . وهذا يعتمد على مجموعة البيانات التي تحتوي على المتغير X بجميـع قيمـه ولكـن الـصياغة تختص بتجميع القيم العشر الأولى. ولجمع كل قيم المتغير X_i يـتم إعـادة كتابـة الـصياغة بحيث تحتوي على الرمز n (عموماً) بدلاً مـن الـرقم الأخـير في المؤشـر، ١٠. أيـضا عنـد عدم احتواء إشارة التجميع على أرقام فإن ذلك يعني أنها من رقم ١ إلى n ويـتم جمـع جميع القيم للمتغير.

تجميع الثابت Summation of a Constant

إذا كان الحرف k يعبر عن رقم ثابت، فـإن مجمـوع k يقـدر بحاصـل ضـرب عـدد مرات التجميع في k. لذا يتم تقدير التعبير الوارد في المعادلة رقم (1.2) بضرب k بالعـدد ٥، عدد الحدود في التجميع، وبالتالي نحصل على الجواب 5k.

$$\sum_{1}^{5} k \tag{1.2}$$

تجميع حاصل ضرب المتغير بثابت Summation of a Constant Times a Variable

في حالة جمع تعبير رياضي مثل $2\ X_i$ حيث 2 عبـارة عـن ثابـت و X_i تعبر عـن المتغير والـتي يمكـن كتابتهـا علـى النحـو التـالي $\sum 2X_i$ ، ولـذلك يمكـن تحويـل الـرقم الثابت خارج عملية التجميع وإعادة كتابتهـا لتـصبح $2\sum X_i$. والتعـبير الأخـير يتطلـب تجميع قيم المتغير ثم ضرب المجموع بالثابت 2.

جبر المجاميع An Algebra of Summation

تتوزع علامة التجميع على التعبير الرياضي بين الأقـواس. وبـذلك يمكـن إعـادة

كتابة التعبير الرياضي $\sum(3X_i + 7)$ بحيث يكون $\sum 3X_i + \sum 7$ ، والذي يمكن اختصاره ليكون $3\sum X_i + 7n$ ، حيث n تمثل عدد الحدود في مؤشر التجميع. ويمكن كتابة التجميع لمربعات المتغيرات على النحو $\sum X_i^2$ وتسميتها مجموع المربعات. ويتم تقديرها بتربيع كل قيمة للمتغير X ثم جمعها ، كما في المعادلة رقم (1.3).

$$\sum X_i^2 = X_1^2 + X_2^2 + ...+ X_n^2. \qquad (1.3)$$

من جهة أخرى فإنه يمكن كتابة مربع المجموع على النحو $\left(\sum X_i\right)^2$ وحسابه عن طريق إيجاد المجموع ثم تربيعه. ومجموع المربعات ومربع المجموع مفهومين مختلفين تماماً ويتم استخدام كلا التعبيرين باستمرار في الصيغ الرياضية.

تمارين Exercises

١- اكتب بالتفصيل المجاميع الممثلة بالتعابير الرياضية التالية :

أ) $\sum X_i$ ب) $\sum X_i^2$

ج) $\sum(Y_i + 4)$ د) $\sum cX_i$

هـ) $\sum f_i X_i^2$ و) $\left(\sum X_i\right)^2$

٢- اكتب التعابير الرياضية التالية باستخدام رمز التجميع والنهايات المناسبة :

أ) $X_1+X_2+X_3+X_4+X_5+X_6$

ب) $X_1^2+X_2^2+X_3^2+X_4^2+X_5^2$

ج) $\left[(X_3+3)+(X_4+3)+(X_5+3)\right]^2$

د) $(X_1-a)^2+(X_2-a)^2+(X_3-a)^2$

٣- إذا كانت قيمة $X_1 = 2$، و $X_2 = 7$ ، $X_3 = -8$ ، $X_4 = -1$ ، $X_5 = 6$ ، $X_6 = 3$ فأوجد القيم الحسابية للفقرات أ ، ب، ج في السؤال الثاني.

الفصل الثاني

تلخيص البيانات
Summarizing Data

عند الحصول على مشاهدات عن واحد أو أكثر من المتغيرات التي تم أخذها من مجتمع أو عينه نكون حصلنا على البيانات. وكما نعلم فإن هناك متغيرات كمية ومتغيرات وصفية ومن ثم لدينا صور مختلفة للبيانات. والمتغيرات الكمية عبارة عن نتيجة لقياس معين أو عملية عدّ معينة وبذلك فإن الأرقام التي تم جمعها كمشاهدات لتلك المتغيرات يتم تصنيفها على شكل بيانات نسبية للمشاهدات التي تم الحصول عليها بطريقة القياس في حين أنه يتم تصنيفها كبيانات فترة للمشاهدات التي تم الحصول عليها بطريقة العدّ.

من جهة أخرى فإن المتغيرات الوصفية تعبّر عن صفات معينة للعينة أو المجتمع مثل العجول، والعجول الفحول، والثيران، والأبقار، والتي تعبّر عن بيانات أسمية كونها عبارة عن أسماء يتم استخدامها من قبل الباحثين للتصنيف على شكل مجموعات بناء على الأسماء، أو تعبّر عن بيانات ترتيبية يتم قياسها بمعيار ترتيبي تتكون من مستويات أو فئات حيث يتم ترتيبها إما تصاعدياً وإما تنازلياً من الأفضل للأسوأ أو العكس.

بالإضافة لما ذكر، فإنه يتم تصنيف البيانات بناءً على مصادرها حيث تصنّف البيانات إلى بيانات أولية عند الحصول عليها عن طريق التجارب، أو استمارات الاستبيان، ومصادر النشر التي حصلت على البيانات مثل التعدادات الزراعية في حين تصنف إلى بيانات ثانوية عند الحصول عليها من مصادر لم تقم بجمع البيانات حتى ولو قامت تلك المصادر بنشر تلك البيانات.

إن الهدف من جمع البيانات يتمثل في استخدامها لاتخاذ قرارات تجاه مشكلة معيّنة من خلال معالجة تلك البيانات وتحليلها إحصائيا ومن ثم الحصول على معلومات عن المشكلة والتي يمكن استخدامها في اتخاذ القرار والذي يتوقع إن يكون ذلك القرار أفضل عند مقارنته بحالة غياب تلك المعلومات.

وبصفة عامة يتم جمع البيانات الإحصائية الخام بطريقة عشوائية كما في الجدول رقم (٢,١) حيث لا يوجد ترتيب معين لتلك البيانات ولذلك فإنه من الصعوبة الحصول على أي قيم معينة أو استدلالات من تلك البيانات في صورتها الخام لذا يتم ترتيب البيانات للمساعدة في تفسيرها.

الجدول رقم (٢,١). إنتاجية نسيج القطن الكتاني (باوند/أيكر) لخمسة و سبعون مزرعة في سهول الأراضي السوداء.

٢٥٥	٣٧٣	٢٤٢	٢٥٧	٣٠٥
٢٨٥	٣٥٨	٢٧٩	٢٦١	٣١٢
٢٨٨	٢٩٧	٣٠٣	٢٩٠	٢٨٣
٢١٥	٢٩٩	٢٧٥	٢٦٠	٢٨٤
٣١١	٢١٧	٢٩٥	٢٦٠	٣١٦
٣١٤	٢٩٠	٢٩٤	٢٨٠	٣٢٦
٢٤٩	٢٥٦	٢٧٦	٢٩٢	٢٨٦
٣٣٣	٢٧٤	٢٥٠	٢٩٥	٣١٨

تابع الجدول رقم (٢،١).

٢٩٢	٢٨٣	٢٧٤	٣٣٦	٣٢٥
٢٩٠	٣٠٩	٢٧٢	٢٩٦	٣٤٦
٢٥٤	٢٣٥	٢٦٨	٢٩٩	٣٥٢
٣٣٤	٢٥١	٣٦٧	٣٠٩	٣١٥
٢٢٨	٢٥٩	٣٥٤	٢٨٣	٣٠٦
٢٣٤	٢٥٨	٣٦٥	٢٨١	٣١٢
٣٤٢	٢٦٨	٢٧٨	٢٩١	٢٨٨

يتم ترتيب البيانات تصاعدياً من أقل قيمه حتى أكبر قيمه ثم حساب المدى بطرح أكبر قيمة من أقل قيمة. في مثال القطن كانت أكبر قيمة ٣٧٣ وأقل قيمة ٢١٥ ولذلك فإن المدى يساوي ١٥٨ رطلاً من الكتان. من ناحية أخرى فإن ترتيب البيانات الجدول رقم (٢،٢) يشير إلى توزيع الوحدات بين أكبر قيمة وأصغر قيمة ومدى ميل تلك البيانات للتجمع حول قيمة معيّنة مثل ٢٩٠ في مثال القطن.

الجدول رقم (٢،٢). إنتاجية نسيج القطن الكتاني (باوند / أيكر) لخمسة و سبعون مزرعة من مزارع سهول الأراضي السوداء مرتبة تصاعدياً.

٢١٥	٢١٧	٢٢٨	٢٣٤	٢٣٥
٢٤٢	٢٤٩	٢٥٠	٢٥١	٢٥٤
٢٥٥	٢٥٦	٢٥٧	٢٥٨	٢٥٩
٢٦٠	٢٦٠	٢٦١	٢٦٨	٢٦٨
٢٧٢	٢٧٤	٢٧٤	٢٧٥	٢٧٦
٢٧٨	٢٧٩	٢٨٠	٢٨١	٢٨٣
٢٨٣	٢٨٣	٢٨٤	٢٨٥	٢٨٦
٢٨٨	٢٨٨	٢٩٠	٢٩٠	٢٩٠

تابع الجدول رقم (٢,٢).

٢٩١	٢٩٢	٢٩٢	٢٩٤	٢٩٥
٢٩٥	٢٩٦	٢٩٧	٢٩٩	٢٩٩
٣٠٣	٣٠٥	٣٠٦	٣٠٩	٣٠٩
٣١١	٣١٢	٣١٢	٣١٤	٣١٥
٣١٦	٣١٨	٣٢٥	٣٢٦	٣٣٣
٣٣٤	٣٣٦	٣٤٢	٣٤٦	٣٥٢
٣٥٤	٣٥٨	٣٦٥	٣٦٧	٣٧٣

يمكن أيضاً تلخيص البيانات في شكل توزيع تكراري كما هو موضح في جدول رقم (٢,٣) حيث يتم إنشاء التوزيع بتقسيم البيانات إلى عدد محدد من الفئات لكل فئة طول معين أو عدد من الوحدات يعبر عن فترة. فبالنسبة لبيانات القطن في جدول رقم (٢,٢) والذي يحتوي على البيانات المرتبة تم تقسيم البيانات لثمان فئات طول كل فئة ٢٠ رطلاً من النسيج الكتاني. ويشير عمود التكرارات إلى عدد المشاهدات التي تقع في الفئة، فمثلاً القيم في الفئة الأولى يمكن أن تصل إلى القيمة ٢٣٥ ولكن لا تساويها. وإذا كانت القيمة ٢٣٥ من بين البيانات فإنها تصبح أحد القيم للفئة الثانية والتي تعتبر القيمة ٢٣٥ حد أدنى لها. من جهة أخرى فإن إجمالي عدد الفئات يجب أن يساوي العدد الإجمالي للبيانات والذي يساوي ٧٥ مزرعة في مثال القطن.

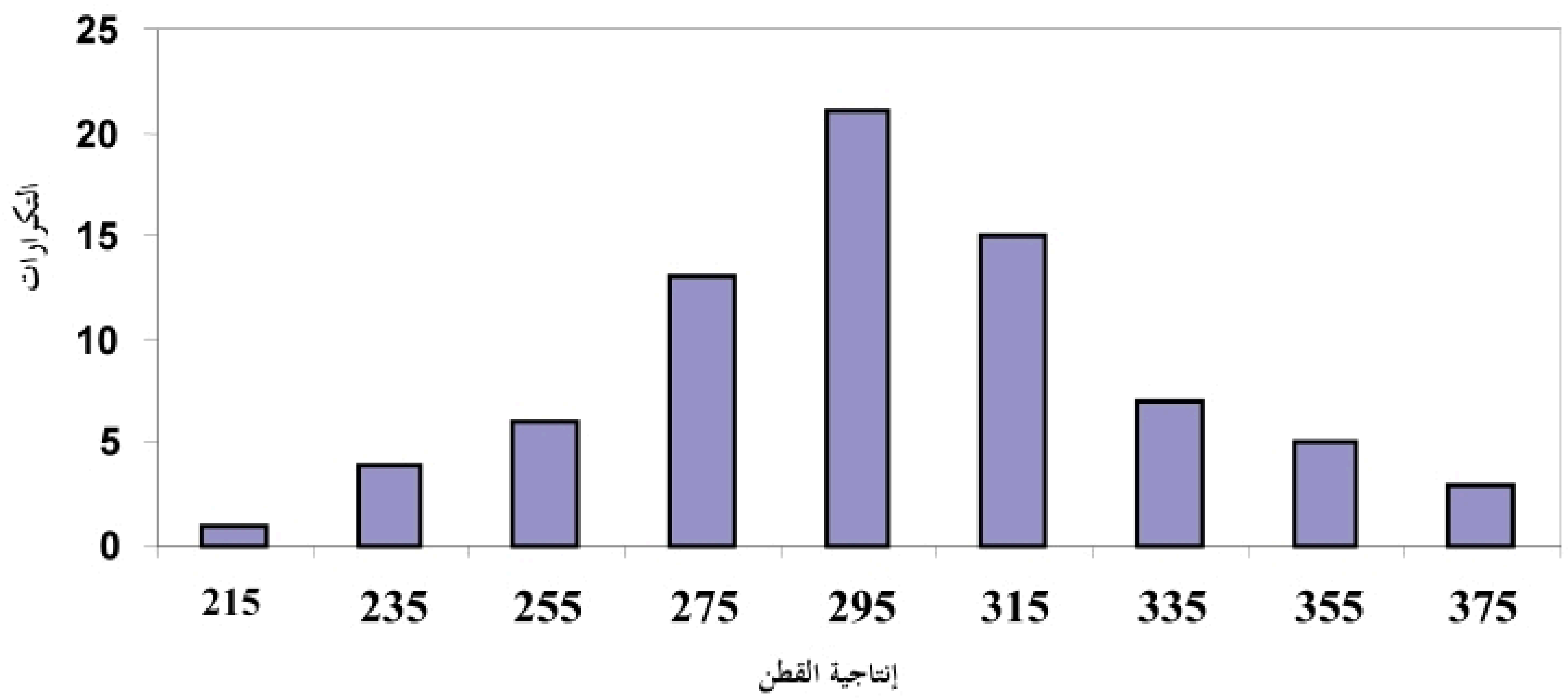
ليس بالضرورة أن يكون طول الفئات متساوي في جميع التوزيعات التكرارية وخاصة عند التعامل مع بيانات الدخل؛ حيث إن استخدام فئات متساوية الطول يؤدي إلى أن بعض الفئات تكون خالية أو تحتوي على مشاهدات قليلة. وفي مثل هذه الحالات يفضل استخدام فئات بطول غير متساوي مما يساهم في تخفيض التوزيع

الجدول رقم (٢،٣). التوزيع التكراري لإنتاجية القطن (باوند/أيكر) لخمسة و سبعون مزرعة من مزارع سهول الأراضي السوداء.

إنتاجية القطن	عدد المزارع
٢١٥ حتى ٢٣٥	٤
٢٣٥ حتى ٢٥٥	٦
٢٥٥ حتى ٢٧٥	١٣
٢٧٥ حتى ٢٩٥	٢١
٢٩٥ حتى ٣١٥	١٥
٣١٥ حتى ٣٣٥	٧
٣٣٥ حتى ٣٥٥	٥
٣٥٥ حتى ٣٧٥	٤
الإجمالي	٧٥

التكراري وبالتالي يساعد في التفسير. في بعض مجموعات البيانات يفضل أن تكون الفئة الأخيرة مفتوحة بدلاً من أن يكون للفئة حد أعلى نظراً لوجود عدد قليل من المشاهدات أكبر من الحد الأعلى المقترح للفئة، وبذلك يمكن كتابة الفئة الأخيرة ٣٥٥ أو أعلى أو أكبر من ٣٥٥. وبالرغم من ذلك إلا أن استخدام هذه الطريقة لها بعض العيوب المتمثلة في عدم القدرة على حساب بعض المتوسطات ومقاييس التشتت مثل المتوسط الحسابي ومتوسط المدى والانحراف المعياري والمدى من التوزيع التكراري.

ويمكن عرض التوزيع التكراري باستخدام الأعمدة البيانية الشكل رقم (٢،١) ويسمى هذا الشكل بالمدرج التكراري Histogram والميزة لاستخدام الرسم في عرض التوزيع التكراري هي إمكانية مشاهدة الشكل الحقيقي لذلك التوزيع ومن ثم يمكن تحديد هل الشكل متماثل أم ملتوي وكذلك مستوى التفرطح بالنسبة للشكل. ويتضح من بيانات القطن في الشكل رقم (٢،١) أنها تقريباً متماثلة وكذلك التوزيع نوعاً ما له



الشكل رقم (٢,١). المدرج التكراري لإنتاجية القطن الكتاني بالباوند/ إيكر مزارع سهول الأراضي السوداء.

قيمة (قيمة عظمى) أي أنه ليس مسطح ولا مدبب.

أما المضلع التكراري Frequency polygon فهو عبارة عن خط بياني يستخدم أحياناً لعرض البيانات حيث يتم عرض التكرارات على المحور الرأسي (الصادي Y) ومراكز الفئات على المحور الأفقي (السيني X) وتوصيل النقاط بخطوط مستقيمة ويتم تحديد مركز فئة تخيّلية (وهمية) قبل بداية البيانات ومركز فئة تخيّلية (وهمية) بعد نهاية البيانات بحيث يكون تكرارها يساوي الصفر وذلك بهدف توصيل المنحنى بالمحور الأفقي ليعطي مدلولاً مناسباً ويتم حساب مراكز الفئات بأخذ متوسط الحد الأعلى للفئة والحد الأدنى. ومثال على ذلك بالنسبة للفئة من ٢١٥ إلى أقل من ٢٣٥ يكون مركزها $(215 + 235)/2 = 225$ وهكذا بالنسبة لجميع الفئات (انظر الجدول رقم ٢,٤).

وتتميز المضلعات التكرارية بسهولة استخدامها مقارنة بالمدرجات التكرارية لعرض شكل التوزيع.ومن المعتاد استخدام المضلعات التكرارية لرسم التكرارات

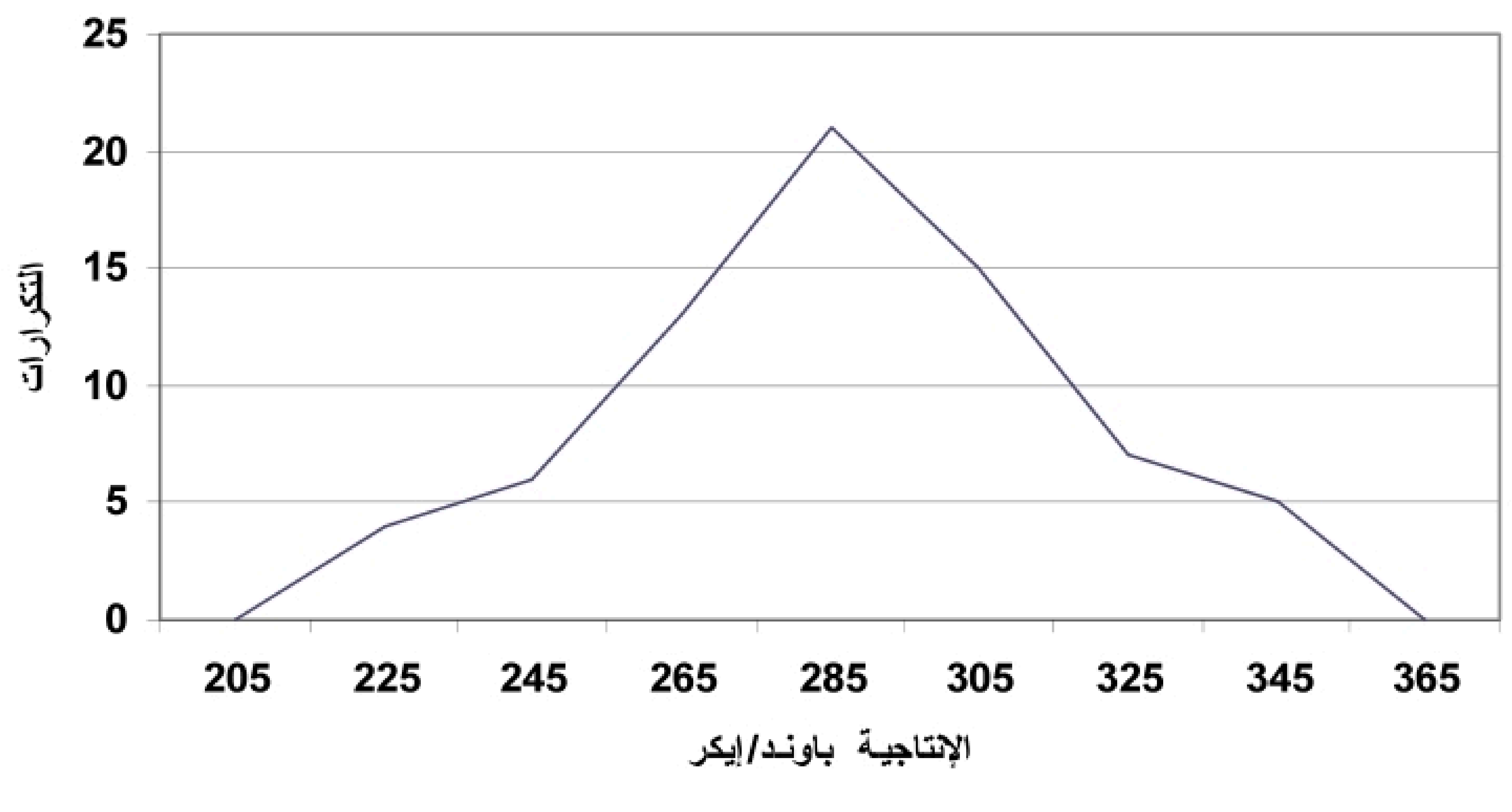
النسبية على نفس المنحنى لمقارنة التوزيعات خاصة إذا كان لدينا عدد مختلف من التكرارات في كل توزيع. يتحدّد التكرار النسبي بقسمة تكرار الفئة على مجموع التكرارات للتوزيع وضربه بـ ١٠٠ لتحويله لنسبة مئوية.ولتوضيح كيفية مقارنة توزيعين يمكن دراسة توزيعات الأجر الأسبوعي لسائقي القاطرات (الشاحنات) وعمال المزرعة كما هو موضح في الجدول رقم (٢,٥). وكما يلاحظ من البيانات بأن أجور العمال الزراعيين بشكل عام أقل ولكن عند رسم تلك البيانات باستخدام المضلع التكراري والموضح بالشكل رقم (٢,٢) نجد صعوبة في المقارنة لصغر العدد بالنسبة للعمال الزراعيين. ويوضح الشكل رقم (٢,٣) التكرارات النسبية للتوزيعين والشكل رقم (٢,٤) المضلعات التكرارية النسبية لهما ، ويلاحظ بشكل عام أن توزيع الأجر بالنسبة لسائقي الشاحنات أكبر من توزيع الأجر بالنسبة للعمال الزراعيين ولكنها تأخذ تقريباً نفس الشكل في المثال السابق.

الجدول رقم (٢,٤). التوزيع التكراري باستخدام مراكز الفئات لبيانات إنتاجية القطن في مزارع سهول الأراضي السوداء.

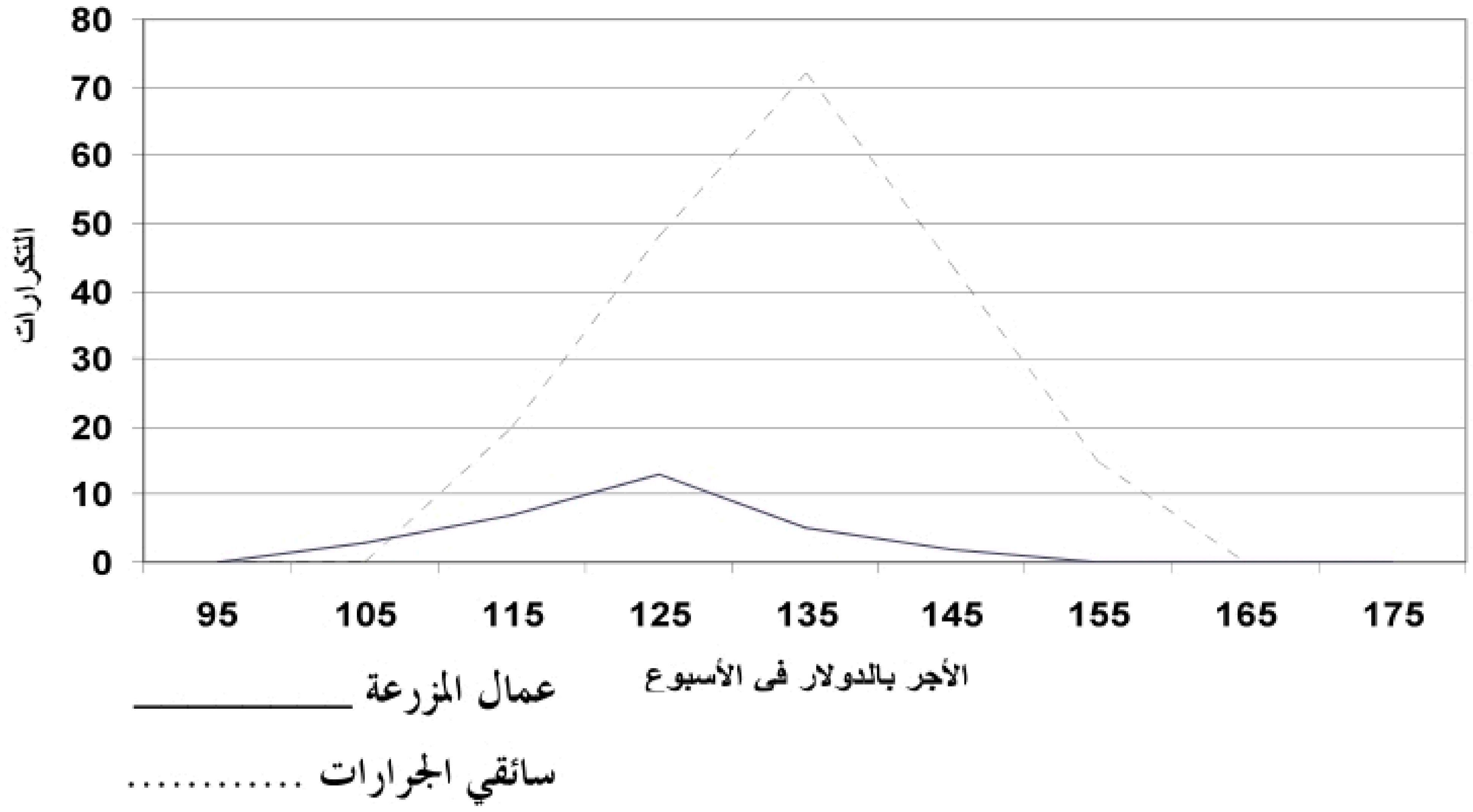
إنتاجية القطن	مراكز الفئات	عدد المزارع
فئة تخيلية	٢٠٥	٠
٢١٥ حتى ٢٣٥	٢٢٥	٤
٢٣٥ حتى ٢٥٥	٢٤٥	٦
٢٥٥ حتى ٢٧٥	٢٦٥	١٣
٢٧٥ حتى ٢٩٥	٢٨٥	٢١
٢٩٥ حتى ٣١٥	٣٠٥	١٥
٣١٥ حتى ٣٣٥	٣٢٥	٧
٣٣٥ حتى ٣٥٥	٣٤٥	٥
فئة تخيلية	٣٦٥	٠
الإجمالي		٧٥

الجدول رقم (٢,٥). التوزيعات التكرارية للأجر الأسبوعي لعمال المزرعة وسائقي الجرارات.

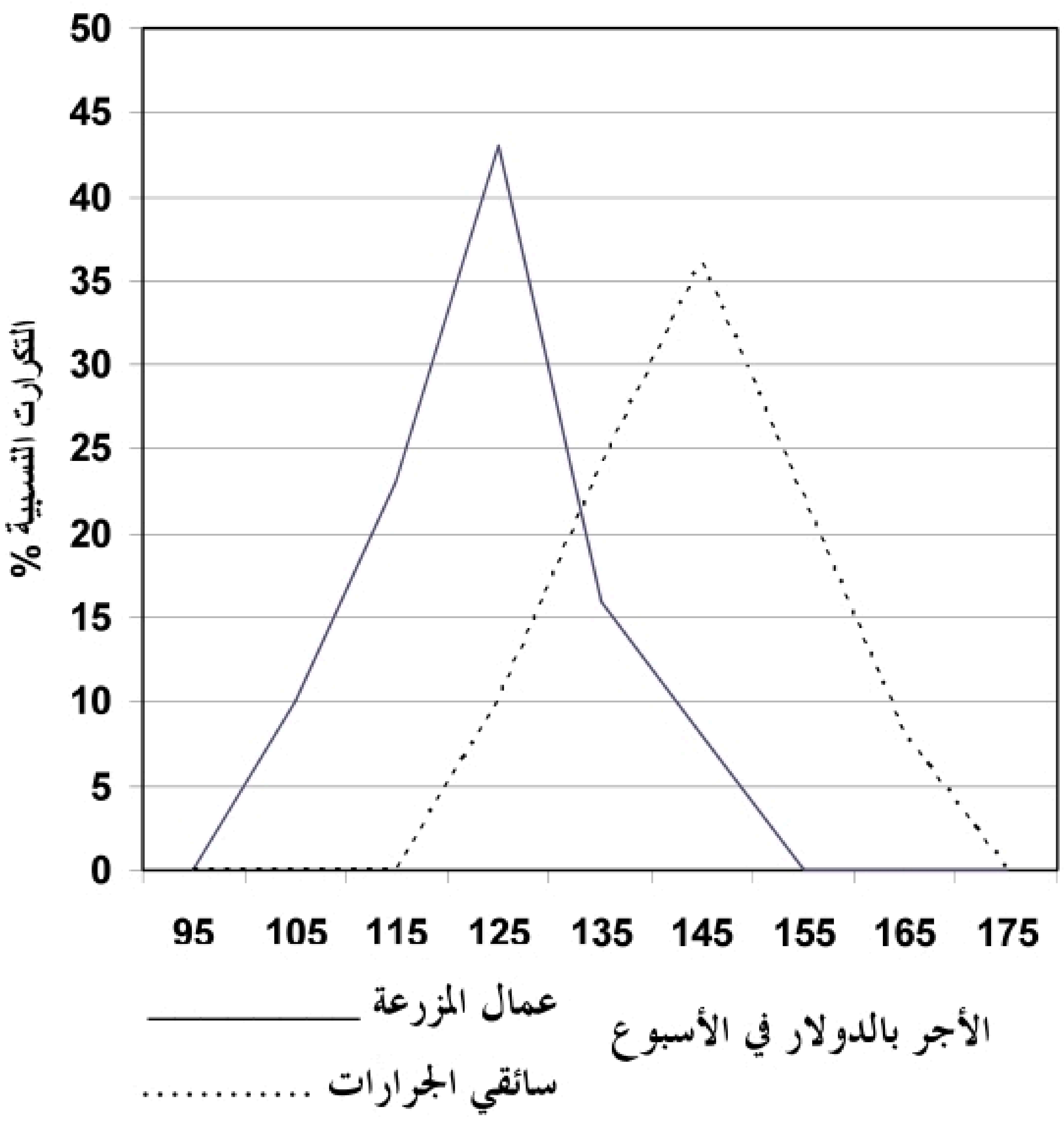
الأجر الأسبوعي	نقاط مراكز الفئة	التكرارات عمال المزرعة	التكرارات سائقي الجرارات	التكرارات النسبية عمال المزرعة	التكرارات النسبية سائقي الجرارات
١٠٠ حتى ١١٠	١٠٥	٣		١٠	٠
١١٠ حتى ١٢٠	١١٥	٧	٠	٢٣	٠
١٢٠ حتى ١٣٠	١٢٥	١٣	٢٠	٤٣	١٠
١٣٠ حتى ١٤٠	١٣٥	٥	٤٨	١٦	٢٤
١٤٠ حتى ١٥٠	١٤٥	٢	٧٢	٨	٣٦
١٥٠ حتى ١٦٠	١٥٥	٠	٤٤	٠	٢٢
١٦٠ حتى ١٧٠	١٦٥	٠	١٦	٠	٨
الإجمالي		٣٠	٢٠٠	١٠٠	١٠٠



الشكل رقم (٢,٢). المضلع التكراري لبيانات إنتاجية القطن في مزارع سهول الأراضي السوداء.



الشكل رقم (٢,٣). المضلع التكراري للأجر الأسبوعي لعمال المزرعة وسائقي الجرارات.



الشكل رقم (٢,٤). المضلع التكراري النسبي للأجر الأسبوعي لعمال المزرعة وسائقي الجرارات.

في المثال السابق تم تحديد عدد الفئات للتوزيع التكراري ولكن بصفة عامة وكقاعدة يتم اختيار القيمة لعدد الفئات في المدى بين ٥ ، ١٥ . حيث إن عدد الفئات أقل من ٥ لايعطي خصائص كافية يمكن من خلالها وصف البيانات وكذلك الحال بالنسبة لعدد فئات أكثر من ١٥ يعتبر كبيراً جداً. في بعض الأحيان يتم استخدام قاعدة سترجس Sturges's rule والموضحة بالمعادلة التالية :

$$K = 1+3.322\,(\log_{10} n)$$

حيث K : عدد الفئات.

n : عدد التكرارات الإجمالي.

وفي الغالب فإن قيمة K لن تكون رقم صحيح وعليه يجب التقريب لأعلى أو أقل رقم صحيح. ويمكن تحديد طول الفئة I بقسمة مدى البيانات (عبارة عن الفرق بين أكبر قيمة وأصغر قيمة) على عدد الفئات K. وبعد تحديد عدد الفئات وطول الفئة يجب التأكد من البيانات ؛ ومدى ملائمة الفئات المختارة لجميع البيانات حيث إنه في بعض الأحيان قد يكون هناك مشاهدة خارج حدود الفئات المحددة ، والاعتبار الآخر الواجب ملاحظته في اختيار طول الفئة هو أن يعرض مركز الفئة نفس البيانات في الفئة. فمثلاً إذا كانت بيانات الأجر تنتهي بـ ٥ دولارات يجب اختيار مركز الفئة بحيث ينتهي بـ ٥ دولارات وذلك بتعديل الفئات ؛ حيث إن قاعدة سترجس تقريبية ويجب أن نقرر في الأخير ما يجب عمله.

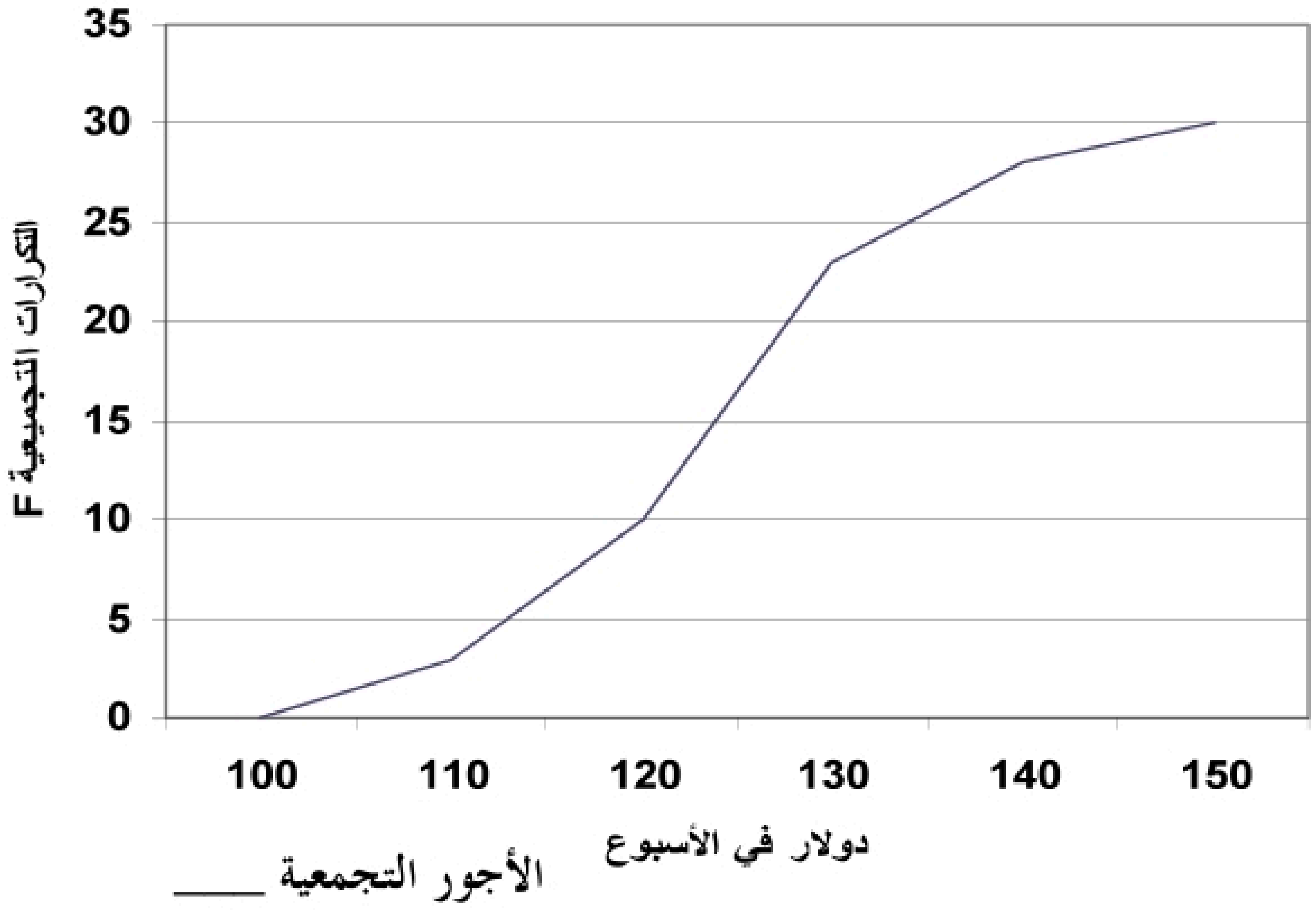
من جهة أخرى ولفهم محتوى البيانات يتم استخدام التوزيعات التكرارية المتجمعة حيث يتم تكوين توزيعات تكرارية صاعدة (أقل من) وتوزيعات تكرارية هابطة (أكثر من). ويتم تكوين التكرارات الصاعدة بجمع التكرارات بْدءاً بأقل فئة

وانتهاءً بأعلى فئة ، وتكتب كلمة أقل من أمام الحد الأعلى للفئة ثم تحسب التكرارات المتجمعة الصاعدة بجمع التكرارات بدْءاً بأقل فئة و انتهاءً بأعلى فئة (الجدول رقم ٢,٦).

الجدول رقم (٢,٦). التوزيع التكراري التجميعي الصاعد لأجور العمال الزراعيين الأسبوعية.

الأجور	التكرارات التجميعية F
أقل من ١٠٠	٠
أقل من ١١٠	٣
أقل من ١٢٠	١٠
أقل من ١٣٠	٢٣
أقل من ١٤٠	٢٨
أقل من ١٥٠	٣٠

ويمكن رسم التكرار المتجمع الصاعد والذي يأخذ شكل حرف S حيث تُمثَّل الفئات (أقل من) على المحور السيني (الأفقي) وتُمثَّل التكرارات المتجمعة الصاعدة على المحور الصادي (الرأسي) (الشكل رقم ٢,٥). وهذا الشكل يساعد في تقسيم التوزيعات التكرارية إلى أجزاء متساوية مثل المئينيات أو العشيريات أو الربيعيات أو أي إحصائيات ترتيبية أخرى ؛ حيث إن المئينيات تقسّم البيانات إلى ١٠٠ جزء متساوٍ وفي الربيعيات تقسم إلى أربعة أجزاء متساوية.

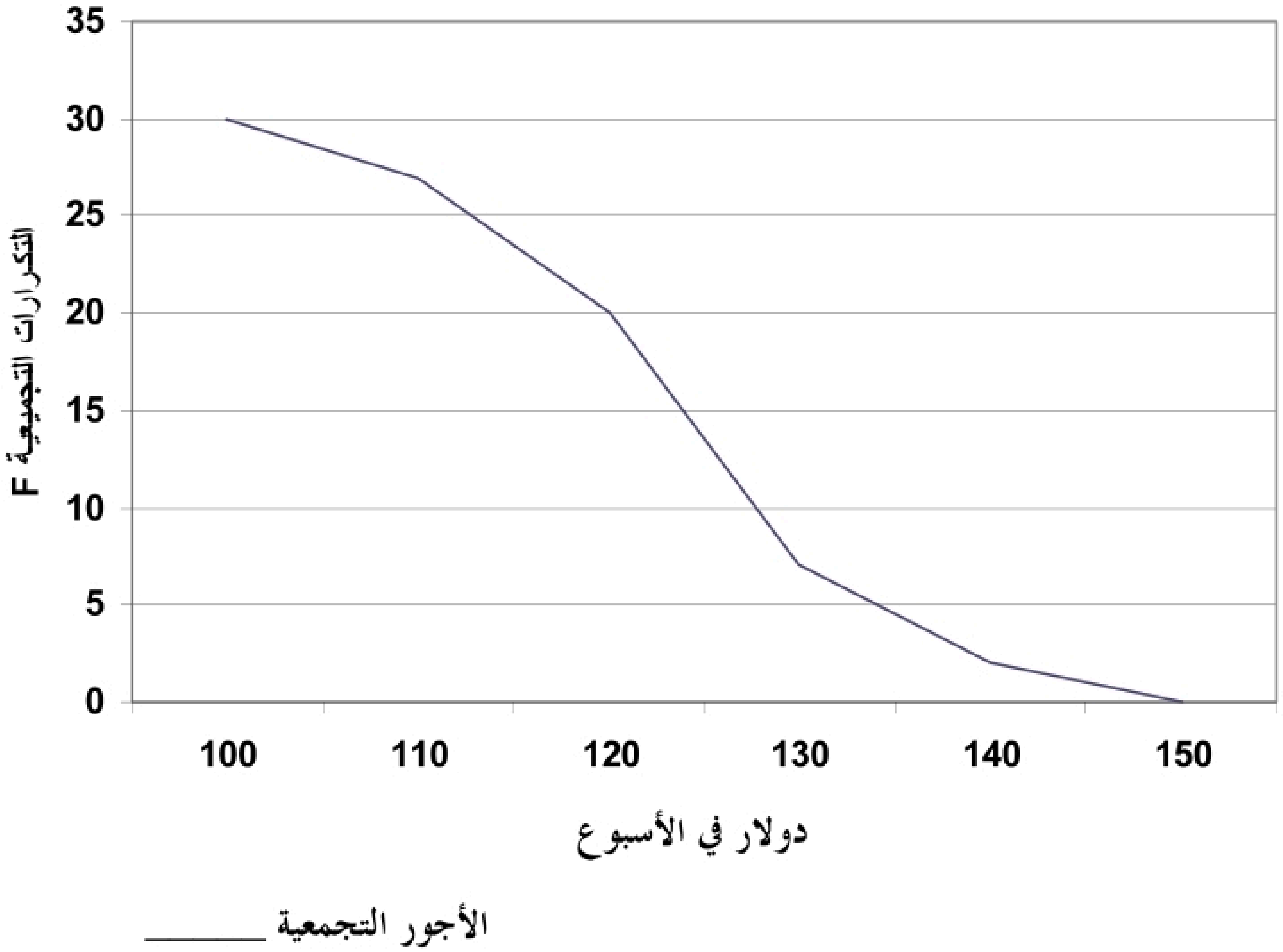


الشكل رقم (٢,٥). المضلع التكراري التجميعي الصاعد لأجور عمال المزرعة.

أما التكرار المتجمع الهابط (أكبر من) يتم البدء بالفئة الأعلى والانتهاء بالفئة الأقل حيث نبدأ بالفئة الأكبر من أقل فئة (الجدول رقم ٢,٧)، ويتم رسم التكرار المتجمع الهابط بحيث نضع الحدود الدنيا للفئات (أكبر من) على المحور السيني والتكرارات المتجمعة على المحور الرأسي ومن ثم نحصل على صورة عكسية للتكرار المتجمع الصاعد (الشكل رقم ٢,٦)، وعند رسم التكرار المتجمع الصاعد و التكرار المتجمع الهابط معاً على رسم واحد فإن الشكلين يتقاطعان في الوسيط.

الجدول رقم (٢,٧). التوزيع التكراري التجميعي الهابط لأجور العمال الزراعيين الأسبوعية.

الأجور	التكرارات التجميعية F
أكثر من ١٠٠	٣٠
أكثر من ١١٠	٢٧
أكثر من ١٢٠	٢٠
أكثر من ١٣٠	٧
أكثر من ١٤٠	٢
أكثر من ١٥٠	٠



الشكل رقم (٢,٦). المنحنى المتجمع الهابط للأجور التجميعية للعمال الزراعيين.

المتوسطات Averages

يعرّف المتوسط بأنه رقم يستخدم للتعبير عن القيمة المركزية لمجموعة من البيانات أو التوزيع. وفي الغالب تستخدم المتوسطات بكثرة مقارنة بالمقاييس الإحصائية الأخرى ويتم حسابها للبيانات الخام (غير المبوبة) و البيانات المبوبة والتي تم تلخيصها في جداول تكرارية. والطريقة الحسابية المستخدمة للنوعين من البيانات سابقاً تختلف في الغالب عن بعضها ؛ نظراً لأننا نحتاج لصيغ رياضية أخرى عند التعامل مع التوزيعات التكرارية.

على الرغم من أن حساب المتوسطات من العينات لا تختلف عنها في المجتمع ولكن بشكل عام يتم استخدام رموز مختلفة للتعبير عن المتوسطات بالنسبة للمجتمع والعينة ، وذلك لمنع اللبس الممكن حدوثه حول نوعية البيانات التي تتم معالجتها. وكما هو معلوم أن العينة مجموعة جزئية من المجتمع في حين أن المجتمع يحتوي جميع البيانات في المجموعة. وفي الغالب تستخدم الأحرف الإغريقية في المعادلات عند التعامل مع بيانات المجتمع.

الوسط الحسابي The Arithmetic Mean

البيانات غير المبوبة Ungrouped Data

يعتبر الوسط الحسابي من أكثر المقاييس استخداماً لقياس النزعة المركزية ويمكن حساب المتوسط ، μ ، للمجتمع بجمع قيم المشاهدات ، X_i ، وقسمتها على عددها N كما في المعادلة رقم (2.1) :

$$\mu = \frac{\sum_{1}^{N} X_i}{N} \qquad (2.1)$$

مثال : أوجد المتوسط الحسابي للبيانات التالية :

7, 3, 2 , 8

$$\sum_{1}^{4} X_i = 7+3+2+8 = 20$$

$$N = 4$$

$$\mu = \frac{\sum_{1}^{N} x_i}{N} = 20/4 = 5$$

وفي حالة أخذ عينة من هذا المجتمع ولنفرض أن حجمها يساوي ٢ يمكن حساب المتوسط الحسابي للعينة باستخدام المعادلة رقم (2.2) :

$$\bar{X} = \frac{\sum_{1}^{n} X_i}{n} \tag{2.2}$$

وتعتمد قيمة متوسط العينة على البيانات التي تم اختيارها من المجتمع سابقاً، ففي حالة اختيار أول مشاهدتين (٧، ٣) فإن المتوسط يكون حاصل جمعها ١٠ مقسوماً على عددها ٢ $\bar{X} = \frac{10}{2} = 5$ والذي يعبر تماماً عن متوسط المجتمع.

ولكن في حالة اختيارات أخرى للعينة من المجتمع فإن متوسط العينة ليس بالضرورة أن يساوي متوسط المجتمع فقد يكون أكبر أو أصغر منه وعليه فإن قيمة متوسط العينة تعتمد على عناصر العينة المختارة من المجتمع ولذلك فإنها من المحتمل أن تساوي أو لا تساوي متوسط المجتمع μ.

وكلما كان حجم العينة المختارة من المجتمع كبير (n) فإن تأثير القيم المتطرفة في المجتمع سيكون قليلاً وبذلك يقترب متوسط العينة من متوسط المجتمع.

في المثال السابق عند اختيار العينة فإن جميع عناصر المجتمع كان لها نفس الوزن

ويساوي ١ ولكن في بعض الأحيان قد نضطر لاستخدام أوزان للعنصر تختلف عن الواحد ويرمز له بـ w_i وفي هذه الحالة فإننا نحسب المتوسط المرجح والذي يمكن حسابه باستخدام المعادلة التالية:

$$\bar{X} = \frac{\sum_{1}^{n} W_i X_i}{\sum_{1}^{n} W_i} \tag{2.3}$$

مثال: البيانات الموضحة في الجدول رقم (٢,٨) توضح الأجر بالساعة وعدد العمال المهاجرين(الموسمين) لمجموعة من محاصيل الخضار والمطلوب حساب المتوسط المرجح.

الجدول رقم (٢,٨). أجر الساعة وعدد العمال الموسميين حسب نوع المحصول.

المحصول	الأجر بالساعة x_i	عدد العمال w_i	wx
خيار	٤,٥	٩٥٠	٤٢٧٥
شمام	٤,٧٥	٦٠٠	٢٨٥٠
بصل	٥,٢٥	١٠٢٠	٥٣٥٥
الإجمالي	_	٢٥٧٠	١٢٤٨٠

يمكن حساب المتوسط المرجح بقسمة المجموع الكلي لحاصل ضرب عدد العمال للمحصول في أجر الساعة لجميع المحاصيل على مجموع العمال باستخدام المعادلة رقم (2.3) $\bar{X} = 12480/2570 = \$4.86$ وبذلك يكون المتوسط المرجح للأجر ٤,٨٦ دولاراً للساعة وهو يكون أدق نظراً لأن عدد العمال غير متساوٍ في كل محصول وعند حساب المتوسط غير المرجح فإنه سيعطى نتيجة مختلفة. وفي الحالات التي تكون

فيها الأوزان كسور عشرية مجموعها يساوي الواحد الصحيح فإن الصيغة الرياضية السابقة يمكن تعديلها بحيث يمكن حذف المقام ؛ نظراً لأن أي قيمة تقسم على الواحد لا تتغير.

خصائص الوسط الحسابي

يوجد خاصيتين للوسط الحسابي هما :

١- مجموع انحرافات القيم عن الوسط الحسابي يساوي صفر.

ويمكن حساب ذلك من المعادلة التالية :

$$x = (X - \overline{X}) \tag{2.4}$$

$$\sum x = 0$$

٢- مجموع مربعات انحرافات القيم عن متوسطها أقل ما يمكن.

فإذا كانت $M = \sum X / n$ عليه فإن $\sum (X - M)^2$ سوف تكون أقل قيمة ممكنة مقارنة بأي قيمة محسوبة بطريقة أخرى لـ M.

البيانات المبوبة Grouped Data

عند تبويب البيانات في شكل توزيعات تكرارية فإن قيم المشاهدات المفردة لن تكون متاحة في الجدول ؛ لأنها على شكل فترات ولذلك يتم التعامل مع التوزيع على شكل فترات ويتم استخدام مركز الفئة كقيمة تقريبية لجميع المشاهدات داخل الفئة. وحيث إن عدد التكرارات يختلف بين الفئات يتم استخدام التكرارات كأوزان في حساب الوسط باستخدام المعادلة التالية :

$$\bar{X} = \frac{\sum_{1}^{n} f_i x_i}{\sum_{1}^{n} f_i} \tag{2.5}$$

حيث f_i تعبر عن التكرار للفئة i

X_i مركز الفئة i

ولنفترض أننا نرغب في حساب متوسط إنتاجية القطن للتوزيع التكراري في الجدول التالي.

الجدول رقم (٢,٩). حساب المتوسط الحسابي لبيانات إنتاجية القطن باستخدام التوزيع التكراري.

إنتاجية القطن	التكرارات، f	مراكز الفئات، X	FX
٢١٥ حتى ٢٣٥	٤	٢٢٥	٩٠٠
٢٣٥ حتى ٢٥٥	٦	٢٤٥	١٤٧٠
٢٥٥ حتى ٢٧٥	١٣	٢٦٥	٣٤٤٥
٢٧٥ حتى ٢٩٥	٢١	٢٨٥	٥٩٨٥
٢٩٥ حتى ٣١٥	١٥	٣٠٥	٤٥٧٥
٣١٥ حتى ٣٣٥	٧	٣٢٥	٢٢٧٥
٣٣٥ حتى ٣٥٥	٥	٣٤٥	١٧٢٥
٣٥٥ حتى ٣٧٥	٤	٣٦٥	١٤٦٠
الإجمالي	٧٥		٢١٨٣٥

في هذه الحالة فإن الوسط الحسابي للتوزيع التكراري يساوي $\overline{X} = 21835/75 = 291.1$ أو ٢٩١ رطلاً للأيكر بينما عند حساب الوسط الحسابي للبيانات الأصلية(انظر الجدول رقم ٢,١) نجد أن المتوسط الحسابي $\overline{X} = 21807/75 = 290.8$ ولذلك فإن استخدام البيانات المبوبة أعطى قيمة أعلى للمتوسط بحوالي ٠,٣ رطل/أيكر.

ومن النتائج السابقة يتضح أن استخدام صيغة التوزيع التكراري بشكل عام قد لا تؤثر على دقة النتائج ؛ لأن عملية إعداد التوزيع التكراري تجعل مركز الفئة تقريب مناسب

للبيانات في الفئة وهذا في الحقيقة يعتبر من الأخبار الجيّدة ؛ نظراً لأن كثير من مصادر البيانات الثانوية في الغالب تكون توزيعات تكرارية ولذلك عند الحاجة لحساب بعض المؤشرات الإحصائية مثل المتوسط يمكن الاعتماد على تلك البيانات بدرجة من الثقة.

متوسط المدى The Midrange

متوسط المدى أو المركز عبارة عن الوسط الحسابي لأعلى قيمة وأقل قيمة للبيانات ولذلك عند ترتيب البيانات تصاعدياً أو تنازلياً فإنه عبارة عن متوسط القيمة الأولى والقيمة الأخيرة.

$$MR = \frac{X_1 + X_n}{2} \tag{2.6}$$

حيث X_1 أقل قيمه ، و X_n أعلى قيمه

ولبيانات إنتاجية القطن الموضحة في جدول رقم (٢,٢) فإن متوسط المدى يساوي :

$$MR = (215 + 373)/2 = 294$$

ويستخدم متوسط المدى في الغالب لحساب المتوسط اليومي لدرجات الحرارة أو متوسط أسعار الأسهم ؛ نظراً لأننا نهتم بأعلى قيمة وأقل قيمة لذلك النوع من البيانات.ولكن بصفة عامة فإن متوسط المدى غير ملائم عند تقدير متوسط عدد السكان ؛ نظراً لأنها تعتمد على قيمتين تتغير من عينة لعينة ولذلك فإن قيمته سوف تختلف كثيراً من عينة لأخرى.

الوسيط The Median

يختلف الوسيط عن المتوسط الحسابي ؛ حيث إن المتوسط الحسابي يعتمد على جميع قيم المشاهدات في حين أن الوسيط يعبّر عن مكان المتوسط. والوسيط لا يتأثر

بالقيم المتطرفة للبيانات ويعتبر ملائماً للاستخدام في الحالات التي نصادف فيها بعض القيم المتطرفة، كما في حالة بيانات الدخل والتعليم.

يحسب الوسيط للبيانات غير المبوبة بترتيب البيانات تصاعدياً أو تنازلياً ويكون الوسيط هو القيمة التي في وسط البيانات إذا كان عدد البيانات فردياً أما إذا كان عدد البيانات زوجياً فإن الوسيط عبارة عن الوسط الحسابي للقيمتين التي في الوسط بعد ترتيب البيانات.

مثال: احسب الوسيط للبيانات التالية ٧، ٣، ٢، ٨

لإيجاد الوسيط ترتب البيانات تصاعدياً ٢، ٣، ٧، ٨ وكون عدد البيانات زوجياً نأخذ متوسط القيم التي في الوسط ٣، ٧ ولذلك فإن الوسيط $5 = \frac{7+3}{2}$.

الجدير بالذكر أن استخدام الوسيط في البيانات القليلة (أقل من ٦ مشاهدات) لا يعطي نتائج دقيقة؛ لأنه يعبّر عن مركز البيانات ولكن في حالة كان عدد المشاهدات أكثر من ٢٠ مشاهدة يمكن الاعتماد على الوسيط كمقياس حيث على الأقل تكون نصف المشاهدات أقل من الوسيط والنصف الآخر أعلى من الوسيط، ويكون الوسيط في حالة التوزيعات المدببة جداً extremely peaked أكثر المقاييس ثقة كتقدير لمتوسط المجتمع.

في حالة البيانات المبوبة يتم حساب الوسيط باستخدام الصيغة الرياضية الموضحة في المعادلة رقم (2.7):

$$\text{Md} = L + \frac{\frac{n}{2} - F}{f} I \qquad (2.7)$$

حيث:

L : الحد الأدنى لفئة الوسيط.

n : مجموع التكرارات.

F : قيمة التكرار المتجمع الصاعد للفئة التي قبل فئة الوسيط.

f : تكرار فئة الوسيط.

I : طول فئة الوسيط.

ولتحديد فئة الوسيط يتم حساب التكرار المتجمع الصاعد ومن ثم حساب رتبة الوسيط والتي تساوي $\frac{n}{2}$ حيث كما نعلم فإن الوسيط عبارة عن القيمة التي تقع في وسط البيانات. ويمكن تحديد الفئة التي تحتوي على الوسيط بعد حساب رتبة الوسيط باستخدام التكرار المتجمع الصاعد؛ حيث إن الحد الأدنى لتلك الفئة هو L وعدد تكرارات تلك الفئة هو f وطول تلك الفئة هو I. أما قيمة F فإنها عبارة عن قيمة التكرار المتجمع الصاعد للفئة التي تسبق الوسيط ومن ثم يتم التعويض بتلك القيم في المعادلة السابقة لحساب قيمة الوسيط.

مثال: لبيانات الأجر لعمال المزرعة الموضحة في الجدول رقم (٢,١٠) احسب قيمة الوسيط.

من تلك البيانات نلاحظ أن مجموع التكرارات يساوي ٣٠ وبذلك يمكن تحديد رتبة الوسيط والتي تساوي $\frac{30}{2} = 15$. ولتحديد فئة الوسيط يمكن مقارنة رتبة الوسيط بقيم التكرارات المتجمعة الصاعدة F في الجدول حيث تقع بين ١٠ و ٢٣، ويشير التكرار المتجمع الصاعد ١٠ إلى أن الفئة ١١٠ إلى أقل من ١٢٠ تحتوي على ٧ قيم تبدأ برقم ٤ وتنتهي بالرقم ١٠ وبصفة مشابهه فإن التكرار المتجمع الصاعد ٢٣ يعني أن الفئة من ١٢٠ إلى أقل من ١٣٠ تحتوي على ثلاث عشرة قيمة تبدأ بالرقم ١١ وتنتهي بالرقم ٢٣ ولذلك فإن الرقم ١٥ يقع في الفئة من ١٢٠ إلى أقل من ١٣٠ والتي تعبّر عن الفئة التي تحتوي على الوسيط أو فئة الوسيط ولذلك فإن القيم هي $L = 120$, $f=13$, $I= 10$, $F=10$

وبالتعويض بتلك القيم في المعادلة رقم (2.7) يمكن حساب الوسيط كالتالي :

$$Md = 120 + \left(\frac{15 - 10}{13}\right)(10) = 120 + 3.8 = 123.8$$

الجدول رقم (٢,١٠). التوزيع التكراري للأجر الأسبوعي للعمال الزراعيين المستخدم في حساب الوسيط.

الأجر اليومي	التكرارات، f	التكرارات التجميعية، F
من ١٠٠ إلى ١١٠	٣	٣
من ١١٠ إلى ١٢٠	٧	١٠
من ١٢٠ إلى ١٣٠	١٣	٢٣
من ١٣٠ إلى ١٤٠	٥	٢٨
من ١٤٠ إلى ١٥٠	٢	٣٠
الإجمالي	٣٠	

المنوال The Mode

يعبّر المنوال عن متوسط آخر حيث يُحدَّد بالقيمة الأكثر تكراراً في مجموعة البيانات. للبيانات غير المبوبة، يتم تحديد المنوال بالبحث حيث يتم ببساطة فحص البيانات قيد الدراسة ومن ثم تحديد القيمة الأكثر تكرار فيها والتي تعبّر عن المنوال.وفي حالة البيانات غير المبوبة قد لا يوجد منوال إذا كانت جميع القيم تظهر لمرة واحدة وقد يوجد أكثر من منوال بالبيانات في حالة تكرار القيم المتشابهة نفس العدد من المرات. وتجدر الإشارة إلى أن المنوال في حالة البيانات غير المبوبة غير حقيقي ؛ نظراً لأن قيمته تتغير في حالة إعادة العينة المسحوبة. ويستخدم المنوال في الحالات التي تتطلب معرفة للحالات المعتادة أو الأكثر شيوعاً مثل برامج التلفزيون المرغوبة، البرامج الأكثر مشاهدة من المشاهدين، أو الاختلافات الأكثر شيوعاً لمحصول الذرة المزروعة ... إلخ.

في حالة البيانات المبوبة أو للتوزيعات التكرارية بشكل عام فإن المنوال عبارة عن القيمة التي تقع عند أعلى نقطة للتوزيع. والمنوال البسيط عبارة عن مركز الفئة التي تحتوي أكثر تكرارات.

ويتضح من الجدول رقم (٢,١٠) أن المنوال البسيط يساوي ١٢٥ ؛ لأن الفئة من (١٢٠ إلى ١٣٠) تحتوي على أعلى تكرار ١٣. ويمكن استخدام الصيغة الرياضية التالية لحساب المنوال معادلة رقم (2.8) .

$$Mode = L + \frac{d_1}{d_1 + d_2} I \qquad (2.8)$$

حيث :

L : الحد الأدنى لفئة المنوال.

d_1 : الفرق الأول : عبارة عن الفرق بين تكرار فئة المنوال وتكرار الفئة السابقة.

d_2 : الفرق الثاني : عبارة عن الفرق بين تكرار فئة المنوال وتكرار الفئة التالية.

I : طول فئة المنوال.

ويمكن تحديد فئة المنوال بالبحث ؛ حيث إنها تحتوي على أعلى تكرار.

وفي حالة المثال السابق الموضح في جدول رقم (٢,١٠) نجد أن الحد الأدنى لفئة المنوال يساوي $L = 120$ والفرق الأول له يساوي٦ ،(١٣- ٧)، والثاني له يساوي ٨ ، (١٣–٥)، وطول فئة المنوال I يساوي ١٠.

وبالتعويض في المعادلة (2.8) نحصل على :

$$Mode = 120 + \left(\frac{6}{6 + 8}\right)(10) = 120 + 4.3 = 124.3$$

وهذه القيمة مساوية تقريباً لقيمة المنوال البسيط والتي تساوي ١٢٥. عندما يكون التوزيع ثنائي القمة Bimodal يفضل تجزئة البيانات إلى مجموعتين وتحليل كل مجموعة على حدة؛ لأن إهمال بعض البيانات وعدم إدخالها في التحليل قد يسبب تكتلاً للبيانات في المجموعتين.

خصائص الوسط والوسيط والمنوال

يمكن استخدام المتوسطات الثلاثة معاً لتحديد التماثل النسبي أو الالتواء للتوزيع. فعندما يكون التوزيع ذو شكل متماثل تماماً فإن قيمة الوسط والوسيط والمنوال تكون متساوية. أما إذا كان التوزيع له ذيل باتجاه اليمين، أو موجب الالتواء، فإن قيمة المتوسط الحسابي تكون الأكبر والمنوال يكون الأصغر قيمة والوسيط يقع في حوالي ثلثي المسافة بين الوسط والمنوال ويكون أقرب للوسط. ويرجع سبب ارتفاع قيمة المتوسط في هذا التوزيع لتأثره بالقيم الشاذة الكبيرة. أما الوسيط فهو يتأثر بمواقع تلك القيم دون أن يتأثر بحجمها. في حالة التوزيع ذي الذيل الأطول باتجاه اليسار، أو سالب الالتواء، تكون قيمة المنوال هي الأكبر وقيمة المتوسط هي الأصغر ويكون الوسيط بين القيمتين بموقع ثلثي المسافة باتجاه الوسط. وعليه فإن معرفة أي قيمتين من تلك القيم الثلاث تمكننا من الحكم على شكل التوزيع هل هو متماثل أو ملتوٍ باتجاه اليمين أو ملتوٍ باتجاه اليسار.

في بعض الأحيان قد يكون التوزيع بدون منوال مثل التوزيعات المفرطحة جداً أو التوزيعات المستطيلة والتوزيعات المتماثلة.

ويمتاز الوسط الحسابي عن المتوسطين الآخرين بإمكانية استخدامه في الحسابات الجبرية ولهذا السبب فإنه من أكثر المقاييس استخداماً، فعلى سبيل المثال يمكن حساب

المتوسط العام لمجموعة من متوسطات العينات بأخذ المتوسط للمتوسطات ثم ترجيحه بعدد المشاهدات في كل عينة.

ومن ناحية أخرى فإن من الصعوبة بل من المستحيل في بعض الأحيان حساب المتوسط عندما يحتوي التوزيع التكراري على فئات مفتوحة في حين أن الوسيط والمنوال لا يتأثر في هذه الحالة حيث يمكن حسابها.

مقاييس التشتت Measures of Dispersion

يعرف التشتت كممثل أو مقياس لمدى صحة المتوسط حيث يبيّن القيمة التي تتوزع بها البيانات سواء في حالة البيانات الخام أو التوزيعات التكرارية حيث تشير القيمة إلى مدى بعد البيانات أو تركزها حول متوسطها. وكلما كان مقدار التشتت للبيانات صغير بالنسبة لمتوسطها دلَّ ذلك على تمثيل المتوسط للبيانات الخام ولذلك يكون لدينا ثقة في استخدام المتوسط كرقم يمثل التوزيع. ولهذا السبب عند استخدام المتوسط كممثل للبيانات يجب أن يضاف له أحد مقاييس التشتت لتمكين المشاهدين من معرفة مدى تمثيل المتوسط للبيانات. فمثلاً في بيانات إنتاجية الذرة إذا كانت الإنتاجية لجميع المزارع تساوي ١٠٠بوشل فإن المتوسط سيكون ١٠٠ سواء أكان الوسط الحسابي أم الوسيط أو المنوال. ويكون التشتت في هذه الحالة مساوٍ للصفر ؛ لأن المتوسط يمثل البيانات تماماً. ولكن في حالة اختلاف أحد القيم عن ١٠٠ فإنه سيكون هناك تشتت أو انتشار في التوزيع. وكلما كان الاختلاف بين بيانات الإنتاجية كبير كلما كان انتشارها في التوزيع كبير وبالتي تقل إمكانية استخدام المتوسط كرقم يمثل مجموعة البيانات.

المدى The Range

المدى ، R ، أحد مقاييس التشتت ويحسب بإيجاد الفرق بين القيمة الأولى والقيمة الأخيرة في البيانات بعد ترتيبها تصاعدياً أو تنازلياً كما في المعادلة رقم (2.9). ويمكن استخدامه مع المتوسط أو الوسيط أو منتصف المدى ولكن لا يمكن استخدامه مع المنوال. وفي الحقيقة ليس للمنوال مقياس تشتت معين مرتبط به ولهذا السبب فإنه أقل المتوسطات استخداماً ، إلا أن الميزة الوحيدة للمنوال هي استخدامه في الحالات غير الرقمية.ويقاس المدى باستخدام المعادلة رقم (2.9).

$$(R) = X_n - X_1 \tag{2.9}$$

وبالإضافة إلى عرض المدى لأعلى قيمة وأقل قيمة في البيانات ، فإنه يستخدم أيضا لمعرفة مدى البيانات نفسها. ويمكن استخدام المدى لقياس تشتت البيانات في حالة العينات الصغيرة ($n \leq 12$) المختارة من التوزيع الطبيعي. ولكن في حالة العينات الكبيرة ($n \geq 30$) فإن المدى يعطي تقدير متحيز لتشتت المجتمع. وفي الحقيقة فإن المدى ليس الاختيار الأول لقياس التشتت ؛ نظراً لاعتماده على قيمتين متطرفتين من مجموعة البيانات ولذلك فإنه يتجه لعدم الاتساق في حالة المعاينة المتكررة من نفس المجتمع.

الانحراف الربيعي The Quartile Deviation

الانحراف الربيعي (QD) أحد مقاييس التشتت ويستخدم مع الوسيط حيث يعبّر عن تشتت البيانات في منتصف وسط التوزيع ويمكن حسابه باستخدام المعادلة رقم (2.10).

$$QD = \frac{Q_3 - Q_1}{2} \tag{2.10}$$

حيث يمثل نصف الفرق بين الربيعي الأول Q_1 والربيعي الثالث Q_3 . وكما نعلم فإنه لحساب الربيعيات تقسم البيانات لأربعة أجزاء متساوية ولعمل ذلك نحتاج لثلاثة مقاييس : الربيعي الأول Q_1 ، الربيعي الثاني أو الوسيط والربيعي الثالث Q_3. وبصفة أخرى فإن الربيعي الأول يعبّر عن الوسيط للنصف الأول من البيانات والربيعي الثالث يعبّر عن الوسيط للنصف الثاني من البيانات ولإيجاد الربيعيات يتم ترتيب البيانات تصاعدياً ثم إيجاد الوسيط Md ، بعد ذلك نوجد الوسيط للنصف الأول من البيانات من X_1 إلى Md ونسميه Q_1 وهو يمثل الربيعي الأول ، وبنفس الطريقة نوجد الوسيط للنصف الثاني من البيانات من الوسيط Md إلى Xn ونسميه Q_3 وهو يمثل الربيعي الثالث. بعد إيجاد الربيعيان نحسب الانحراف الربيعي ونستخدمه مع الوسيط.

مثال: البيانات التالية توضح عدد سنوات التعليم لعينة صغيرة من البالغين في الريف Rural ٨ ، ١٢ ، ٦ ، ١٤ ، ١٠

والمطلوب حساب الانحراف الربيعي.

الحل:

أولاً يتم ترتيب البيانات تصاعدياً ٦ ، ٨ ، ١٠ ، ١٢ ، ١٤ ثم نوجد الوسيط والذي يساوي ١٠ والربيعي الأول Q_1= ٧ والربيعي الثالث Q_3 = ١٣ ولذلك فإن الانحراف الربيعي يساوي :

$$QD = (13 - 7) / 2 = 3$$

ويتشابه الانحراف الربيعي إلى حد ما مع المدى ؛ نظراً لأنه يقيس الفرق بين قيمتين ولكن تلك القيمتين في منتصف وسط التوزيع. ونظراً لأن جزءاً من البيانات

الواقع بين الربيعي الثالث ونهاية البيانات لا يدخل في حساب الانحراف الربيعي فإن استخدام هذا المقياس يعطي نتائج غير دقيقه عندما يكون هناك تشتت كبير في هذا الجزء من البيانات. من جهة أخري، فإنه في حالة حساب المتوسط للبيانات ذات النهايات المفتوحة فإن الانحراف الربيعي يعتبر المقياس المناسب لتشتت هذه البيانات.

في حالة التوزيعات التكرارية (البيانات المبوبة) يمكن حساب الربيعي الأول Q_1، الربيعي الثالث Q_3 باستخدام صيغ رياضية مشابهة للتي استخدمت لحساب الوسيط والموضحة في المعادلات رقم (2.11) و (2.12).

$$Q_1 = L + \frac{\frac{n}{4} - F}{f} I \tag{2.11}$$

$$Q_3 = L + \frac{\frac{3n}{4} - F}{f} I \tag{2.12}$$

وهي نفس الصيغة المستخدمة لحساب الوسيط ماعدا أنه يتم تحديد الفئة التي تحتوي على الربيعي الأول بمقارنه $\frac{n}{4}$ مع التكرار المتجمع الصاعد F ، وتحديد الفئة التي تحتوي على الربيعي الثالث بمقارنة $\frac{3n}{4}$ مع التكرار المتجمع الصاعد F ، وبتحديد الفئات لكل ربيعي فإن L ، f ، I يمكن إيجادها و F هي عبارة عن التكرار المتجمع الصاعد للفئة السابقة لها.

الانحراف المعياري The Standard Deviation

يستخدم الانحراف المعياري كمقياس للتشتت مع الوسط الحسابي . وتعتمد قيمته على جميع المشاهدات في البيانات . ويسمى في بعض الأحيان بجذر متوسط مربعات الانحرافات ؛ نظراً لأنه يتم حسابه بأخذ الجذر التربيعي للوسط الحسابي لمربعات الانحرافات عن المتوسط.

البيانات غير المبوبة *Ungrouped Data*

يتم حساب الانحراف المعياري للبيانات غير المبوبة للمجتمع والذي يرمز له بالرمز σ (سيجما) باستخدام المعادلة رقم (2.13) :

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}} \tag{2.13}$$

ويكون التباين σ^2 هو مربع الانحراف المعياري. ولكن من الناحية الرياضية يتم إيجاد التباين أولاً ثم حساب الانحراف المعياري بأخذ الجذر التربيعي للتباين إذا كان الهدف حساب الانحراف المعياري. ويمتاز الانحراف المعياري بأنه أكثر مقاييس التشتت استخداماً ؛ نظراً لأن الوسط الحسابي أكثر مقاييس النزعة المركزية استخداماً والذي يستخدم معه الانحراف المعياري إضافة إلى سهولة حساب التباين والتعامل معه رياضياً واستخدام تلك الحسابات في عمليات حسابية أخرى.

ويرمز لتباين العينة بـ S^2 ، وهو عبارة عن تقدير لتباين المجتمع σ^2 ، ويحسب من بيانات العينة المسحوبة من المجتمع. وكما هو معلوم فإنه في المعاينة المتكررة فإن تباين العينة يكون متحيز وأقل من تباين المجتمع بقيمة ثابته تساوي $(n-1)/n$. وبناء

على ذلك ، فإن معظم الإحصائيين عدّل معادلة تباين العينة بالقسمة على $(n-1)$ بدلاً من n حتى يتم التخلص من أثر التحيز ويحسب تباين العينة باستخدام المعادلة رقم (2.14) والتي يمكن صياغتها كالتالي :

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}} \qquad (2.14)$$

مثال :

افترض أنه في إحدى مؤسسات الأعمال الزراعية الكبيرة تم اختيار سجل الغياب بسبب المرض في منتصف السنة لعدد من الموظفين وتسجيل عدد أيام غيابهم عن العمل فكانت على النحو التالي : ٧ ، ١٤ ، ٨ ، ٥ ، ١٥ ، ١١

المطلوب حساب المتوسط الحسابي والانحراف المعياري لهذه البيانات .

لحساب الانحراف المعياري للعينة يتم :

١- حساب المتوسط الحسابي للعينة $\bar{X}$ بإيجاد حاصل جمع البيانات $\sum x$ ثم قسمتها على عددها :

$$\bar{X} = \frac{\sum x}{n} = \frac{7+14+8+5+15+11}{6} = \frac{60}{6} = 10$$

٢- نحسب الانحراف للبيانات عن المتوسط $(X - \bar{X})$ لجميع المشاهدات مع ملاحظة أن مجموعها يساوي صفر :

$$\sum(X - \bar{X}) = 0$$

٣- نحسب مربع الانحراف $(X - \bar{X})^2$ لجميع المشاهدات ثم نوجد مجموعها $\sum(X - \bar{X})^2$ ويوضح الجدول رقم (٢,١١) نتائج تلك الحسابات.

الجدول رقم (٢,١١). بيانات توضح طريقة حساب الانحراف المعياري للعينة.

عدد أيام الغياب x	$(X - \overline{X})$	$(X - \overline{X})^2$
٧	٣ -	٩
١٤	٤	١٦
٨	٢ -	٤
٥	٥ -	٢٥
١٥	٥	٢٥
١١	١	١
٦٠	٠	٨٠

ولحساب التباين للعينة يتم التعويض في المعادلة التالية :

$$S^2 = \frac{\sum (x - \overline{x})^2}{n - 1} = \frac{80}{6 - 1} = \frac{80}{5} = 16$$

ثم إيجاد الانحراف المعياري $S = \sqrt{S^2} = \sqrt{16} = 4$

وكما هو ملاحظ سهولة العمليات الحسابية في هذا المثال ؛ نظراً لأن قيم الانحراف عن المتوسط عبارة عن أعداد صحيحة حيث تكون الحسابات أكثر تعقيدا في حالة وجود قيم كسرية.

وعليه فإنه يمكن استخدام صيغة رياضية أخرى لحساب الانحراف المعياري (المعادلة رقم (2.15)) بحيث لا نستخدم الانحراف عن المتوسط في الحسابات .

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1}} \qquad (2.15)$$

ولحساب الانحراف المعياري S في المثال السابق يجب حساب قيم X^2 للمتغير كالتالي :

X^2	X
٤٩	٧
١٩٦	١٤
٦٤	٨
٢٥	٥
٢٢٥	١٥
١٢١	١١
٦٨٠	٦٠

وبالتعويض في المعادلة (2.15) بقيم $n \;,\; \sum x \,,\, \sum x^2$ ثم حل المعادلة:

$$S = \sqrt{\frac{680 - \frac{(60)^2}{6}}{6-1}} = 4$$

وهي نفس النتيجة التي تم الحصول عليها باستخدام الصيغة الرياضية السابقة (2.14).

البيانات المبوبة Grouped Data

لحساب الانحراف المعياري للبيانات المبوبة في شكل التوزيعات التكرارية يتم استخدام التكرارات f في الحساب كأوزان. لذا يتم تعديل جميع الصيغ بحيث تشمل الأوزان. وتوضح المعادلات رقم (2.16) و (2.17) الصيغ الرياضية لحساب الانحراف المعياري لبيانات العينة المبوبة على شكل توزيعات تكرارية حيث يتم استخدام الفروق في المعادلة رقم (2.16) و التجميع في المعادلة رقم (2.17). ولتوضيح كيفية حساب S بمثال نعود مرة أخرى لبيانات أجور العمال الزراعيين الموضحة في الجدول رقم (٢,١٢). ويلاحظ أن الجدول إضافة لبيانات التوزيع التكراري الأساسية يحتوي على أعمدة

تشمل X ، fX ، X^2 و fX^2 والتي نحتاج لها للتعويض في المعادلات.

$$s = \sqrt{\frac{\sum f\ (X - \bar{X})^2}{\sum f - 1}} \tag{2.16}$$

$$s = \sqrt{\frac{\sum f\ X^2 - \frac{\left(\sum fX\right)^2}{\sum f}}{\sum f - 1}} \tag{2.17}$$

ويجب أن نضرب كل قيمة لـ X بتكرارها f ثم نقوم بتجميع تلك القيم للحصول على المجموع $\sum fX$ ، وأيضا نربع قيم X ثم نضربها بالتكرارات للحصول على $f X^2$ ثم نقوم بتجميع تلك القيم للحصول على المجموع $\sum f X^2$. والقيمة الأخيرة التي نحتاجها هي $\sum f$ والتي يتم الحصول عليها بتجميع عمود التكرارات.
وبالتعويض عن القيم في المعادلة أعلاه نحصل على :

$$s = \left[\left(461950 - 3710^2 / 30\right)/\left(30 - 1\right)\right]^{0.5} = \left[108.5\right]^{0.5} = 10.4$$

الجدول رقم (٢,١٢). حسابات الانحراف المعياري للتوزيع التكراري لأجور المزرعة.

الأجور الأسبوعية	التكرارات f	مراكز الفئات X	X^2	fX	fX^2
١٠٠ حتى ١١٠	٣	١٠٥	١١٠٢٥	٣١٥	٣٣٠٧٥
١١٠ حتى ١٢٠	٧	١١٥	١٣٢٢٥	٨٠٥	٩٢٥٧٥
١٢٠ حتى ١٣٠	١٣	١٢٥	١٥٦٢٥	١٦٢٥	٢٠٣١٢٥
١٣٠ حتى ١٤٠	٥	١٣٥	١٨٢٢٥	٦٧٥	٩١١٢٥
١٤٠ حتى ١٥٠	٢	١٤٥	٢١٠٢٥	٢٩٠	٤٢٠٥٠
المجموع	٣٠	-	-	٣٧١٠	٤٦١٩٥٠

خصائص الانحراف المعياري S و المتوسط $\bar{X}$

يوجد خاصيتان للمتوسط والانحراف المعياري جديرة بالملاحظة خاصة عندما يجب أن نعمل تعديل للبيانات. فمثلاً لو رغبنا إضافة رقم ثابت لكل عنصر في مجموعة البيانات ماذا سيحدث لقيمة S ؟ وفقاً للخاصية الأولى فإن قيمة المتوسط تزيد بمقدار مساوٍ لقيمة الثابت في حين أن قيمة الانحراف المعياري لاتتغير بإضافة رقم ثابت لكل قيمة[١]. الخاصية الثانية تنص على أنه في حالة ضرب كل قيمة من القيم برقم ثابت فإن المتوسط يساوي حاصل ضرب القيمة لذلك الرقم في المتوسط والانحراف المعياري يساوي حاصل ضرب القيمة المطلقة للرقم الثابت في الانحراف المعياري[٢] في حين أن التباين يساوي حاصل ضرب مربع ذلك الرقم في التباين. ونجد أن تلك الخصائص مفيدة ؛ نظراً لعدم الحاجة لحسابات حقيقية للمتوسط أو التباين للبيانات المعدلة. حيث يتم استخدام الخاصية للمتوسط أو الانحراف المعياري التي تم حسابها من البيانات الأصلية للحصول على القيم المطلوبة للبيانات المعدلة.

ويمكن أيضاً معايرة البيانات بتحويل المتوسط لقيمة تساوي الصفر والانحراف المعياري لقيمة تساوي الواحد. وللحصول على متوسط يساوي الصفر نطرح المتوسط من كل قيمة من القيم في مجموعة البيانات على حدة. ولذلك يكون متوسط البيانات المعدلة مساوٍ للصفر ولكن الانحراف المعياري كما هو دون تغيير. ولتغيير الانحراف المعياري ليصبح مساوٍ للواحد نقسم كل قيمة من البيانات على الانحراف المعياري S (نضربها بـ 1/S). وهذه العملية الحسابية تقسم المتوسط على S ، ولكن قسمة الصفر على رقم ما عدا الصفر نحصل على قيمة تساوي الصفر. وبشكل عام يتم التعبير عن البيانات المعيارية للمجتمع بالرمز z والذي يساوي $Z_i = (X_i - \mu)/\sigma$.

معامل الاختلاف Coefficient of Variation

عندما نرغب في مقياس للاختلافات الموجودة في مجموعة البيانات فإننا نستخدم الانحراف المعياري مع المتوسط. ويتم تفضيل الانحراف المعياري ؛ لأن له نفس وحدات المتوسط بخلاف التباين الذي يعبر عن وحدات مربعة. ولكن في الغالب يتم حساب معامل الاختلاف لتحديد الاختلاف النسبي في مجموعة البيانات.وهو عبارة عن نسبة الانحراف المعياري إلى المتوسط مضروباً في ١٠٠ ، أي $V = (S / \bar{X})(100)$. ولذلك فإن معامل الاختلاف يعبر عن مدى كبر الانحراف المعياري مقارنة بالمتوسط على شكل نسبي. فإذا كانت قيمة V تساوي ١٠٠ فإنه يشير إلى تساوي الانحراف المعياري والمتوسط. وفي هذه الحالة فإن الاختلاف في البيانات كبير والمتوسط في هذه الحالة ليس المقياس المناسب لمركز التوزيع. وكلما كانت قيمة معامل الاختلاف V صغيرة كان المتوسط ممثل جيد لمجموعة البيانات. وكقاعدة فإننا نستخدم الحذر عندما تكون قيمة معامل الاختلاف V أعلى من ٥٠ ٪ حيال عرض البيانات باستخدام المتوسط.

من جهة أخرى لا يمكن مقارنة الانحرافات المعيارية لتوزيعين ؛ نظراً لارتباط قيم S بقيم المتوسطات لتلك التوزيعات.أي ، أنه لا يوجد تفسير للانحراف المعياري بدون المتوسط (ربما يستثنى من ذلك بعض الحالات غير المعتادة عندما تكون المتوسطات متساوية). وبناء على ذلك يتم مقارنة قيم V للتوزيعات بدلاً من ذلك. وكلما كانت قيمة V للتوزيع قليلة كان التوزيع أقل اختلاف.

ملاحظات ختامية

(١) إذا كان $s^2 = \sum \left(X - \bar{X} \right)^2 / (n-1)$ وتم إضافة ثابت k لـ X بحيث يكون لدينا X+k فإن :

$\overline{X} = \sum (X+k)/n = \sum X/n + nk/n = \overline{X} + k$ وبنفس الطريقة يمكن التعويض في التباين كالتالي :

$$s^2 = \sum\left[(X+k) - (\overline{X}+k)\right]^2/(n-1) = \sum\left[X+k-\overline{X}-k\right]^2/(n-1) = \sum\left(X-\overline{X}\right)^2/(n-1)$$

وهي نفس الصيغة الأساسية للتباين.

(٢) إذا كان لدينا kX فإن المتوسط يكون :

$\overline{X} = \sum kX/n = k\sum X/n = k\overline{X}$ والتباين يمكن التعبير عنه بـ :

$$s^2 = \sum\left[kX - k\overline{X}\right]^2/(n-1) = \sum[k^2X^2 - 2kX\overline{X} + k^2\overline{X}^2]/(n-1) =$$

$$\sum\left[k^2\left(X^2 - 2X\overline{X} + \overline{X}^2\right)\right]/(n-1) = k^2\left[\sum X^2 - 2X\overline{X} + \overline{X}^2\right]/(n-1) = k^2\sum(X-\overline{X})^2/(n-1) = k^2S^2$$

وبذلك يكون الانحراف المعياري مساوي kS .

تمارين Exercises

١- لبيانات العينة التالية (-١٠ ، ٢ ، ٣ ، ٢ ، -٤ ، ٢ ، ٥) أوجد ما يلي :

أ) المتوسط الحسابي.

ب) الوسيط.

ج) المنوال.

د) متوسط المدى.

هـ) المدى.

و) الانحراف المعياري.

٢- للبيانات في التمرين ١ احسب الربيعي الأول والثالث ثم الانحراف الربيعي.

٣- إذا أعطيت مجموعة الأرقام التالية : ٨ ، ٥ ، ٢ ، ٦ ، ٤ ، ٥ المطلوب :

أ) أوجد المتوسط والوسيط والمنوال لهذا المجتمع.

ب) أوجد التباين والانحراف المعياري.

ج) احسب المدى ومتوسط المدى.

٤- البيانات التالية توضح الراتب السنوي عند التعيين لعدد ١٦ خريج حديث من كلية الزراعة :

٢٦٥٠٠دولار	١٩٩٠٠دولار	٣١٢٠٠دولار	٣١٤٠٠دولار
٢٠٤٠٠دولار	٢١٤٠٠دولار	٢١٨٠٠دولار	٢٥٥٠٠دولار
٢٤٦٠٠دولار	٢٢٦٠٠دولار	٢٤٨٠٠دولار	٢٧٠٠٠دولار
٢٣٦٠٠دولار	٢٨٤٠٠دولار	٢٣٤٠٠دولار	٢٩١٠٠دولار

أ) احسب المتوسط الحسابي والوسيط لهذه البيانات.

ب) أستخدم فئات بطول ٢٥٠٠دولار لإيجاد التوزيع التكراري للبيانات. ثم أوجد المدرج التكراري لهذا التوزيع وارسم المضلع التكراري بتوصيل مراكز الفئات (إبداء بالفترة ١٩٠٠٠دولار-٢١٥٠٠دولار).

ج) أوجد منحى التكرار النسبي المتجمع الصاعد ثم ارسمه. كم نسبة الخريجين الذين رواتبهم أقل من ٢٤٠٠٠دولار؟ أكثر من ٢٩٠٠٠دولار.

د) استخدم التوزيع التكراري الذي تحصلت عليه في الجزء ب لحساب المتوسط الحسابي والانحراف المعياري. كم نسبة الرواتب الواقعة في المدى واحد انحراف معياري للمتوسط؟ اثنين انحراف معياري ؟

هـ) كوّن التوزيع التكراري للجزء ب ثم احسب الوسيط ، الانحراف الربيعي والمنوال.

٥- البيانات التالية توضح التوزيع التكراري لكمية السعرات الحرارية في حليب يحتوي على نسبة دهون ٤٪ لعينة مأخوذة من ١٠٠ بقرة :

السعرات الحرارية	التكرار
من ٩٠ إلى أقل من ١٠٠	١٢
من ١٠٠ إلى أقل من ١١٠	٥٥
من ١١٠ إلى أقل من ١٢٠	٢٥
من ١٢٠ إلى أقل من ١٣٠	٨
الإجمالي	١٠٠

أوجد:

أ) المتوسط.

ب) الانحراف المعياري.

ج) الوسيط

د) المنوال.

هـ) ارسم المدرج التكراري للتوزيع التكراري ثم علق على تماثله.

٦- كيف تكون العلاقة بين المتوسط والوسيط والمنوال في حالة التوزيعات التكرارية المتماثلة؟ وكيف تكون تلك العلاقة في حالة التوزيعات التكرارية الملتوية لليمين؟ الملتوية لليسار؟ ارسم عدة توزيعات لتوضيح إجابتك.

استخدام برنامج أكسل أو أي برنامج حاسوبي مناسب لحل التمرين التالي:

٧- البيانات التالية توضح الوزن المكتسب خلال ٢٠٥ أيام لعدد ٢٤ من عجول التغذية نوع البرانقس brangus السوداء .

٤٨٢	٥٠٥	٤٦٧	٥٢١
٥٥٠	٥٣٤	٤٨٥	٥٤٢
٤٧٠	٥١١	٤٩٠	٥١٧
٥٤٥	٤٧٦	٤٨٧	٥٣٠
٤٩٦	٥٠٤	٥٥٨	٥٣٦
٤٦٣	٤٧٠	٥٥٣	٥١٢

احسب :

أ) المتوسط والانحراف المعياري للبيانات الخام باستخدام الأدوات ثم تحليل البيانات ثم أمر الإحصاء الوصفي.

ب) التوزيع التكراري بفترات مداها ٢٠ رطلاً مبتدئاً بـ ٤٦٠ ومنتهياً بـ ٥٦٠ باستخدام أدوات ، تحليل البيانات ، أمر المدرج التكراري.

ملاحظة: في العمود القريب أنشئ القيم ٤٦٠ ، ٤٨٠ ، ٥٠٠ ، ٥٢٠ ، ٥٦٠ ، قبل الضغط على أدوات في قائمة الأوامر. أدخل مدى هذا العمود في جدول تعريف المدرج التكراري الذي يظهر لك. أيضاً أشر على الصندوق الذي أمام الرسم ؛ وذلك لرسم البيانات.

الفصل الثالث

الاحتمــــال

Probability

يتطلب الاستدلال الإحصائي فهم نظرية الاحتمالات لربط خصائص العينة بالمجتمع الذي سحبت منه. لذلك فإننا نحتاج معرفة نظرية الاحتمالات لإجراء عمليات الاستدلال الإحصائي Statistical Inference.

وبشكل عام فإنه يتم استخدام نظرية الاحتمالات لرسم خلاصة حول مكونات العينة بناء على نموذج رياضي للمجتمع. ولكن يتم استخدام الاستدلال الإحصائي لرسم خلاصة حول النموذج الرياضي للمجتمع بناء على بعض الإحصاءات المحسوبة من العينة.

وتتضمن دراسة الاحتمالات ثلاثة أنواع من المسائل:

١- تعريف وتفسير معنى الاحتمالات.

٢- استخدام الاحتمالات المعروفة لحساب الاحتمالات الأخرى.

٣- الحصول على الاحتمالات العددية.

وسيتم التطرق للنوعين الأول والثاني في حين سيتم التطرق للنوع الثالث عند دارسة موضوع التقدير. أيضاً، ستتم مناقشة مواضيع نظرية الاحتمال الأكثر فائدة في

فهم الاستدلال الإحصائي وسوف يتم الاهتمام بمعنى الاحتمالات ووضع افتراضات باستخدام قواعد الاحتمالات. بعد عرض بعض طرق العد سيتم التطرق لنظرية الاحتمالات بحيث تتضمن موضوعات عن نظرية المجموعة، قواعد حساب الاحتمالات، مراجعة الاحتمالات والتوقع الرياضي.

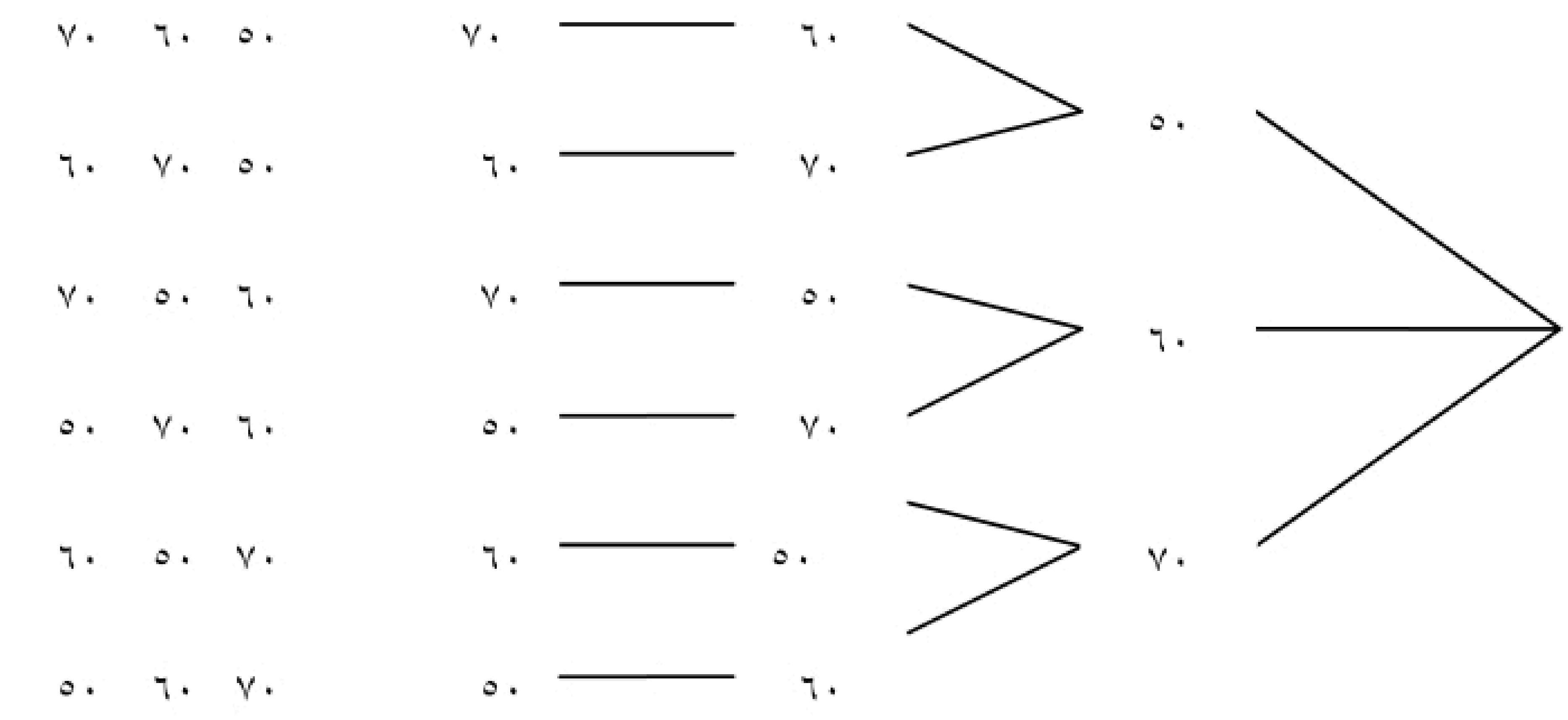
طرق العد Methods of Counting

لقد تعلمنا العدّ مبكراً في الحياة، وعند القيام بالعد يتم ترتيب الوحدات أو وضعها في قائمة. وتشتمل طريقة العد المهمة في نظرية الاحتمالات على اختيار عدد الطرق التي يمكن أن نرتب بها مجموعة البيانات.

التباديل Permutations

تعرف التباديل بأنها ترتيب لمجموعة من الأشياء مع مراعاة الترتيب. على سبيل المثال نفترض أن بائع الآلات يرغب في وضع ثلاث حراثات جديدة في صف لعرضها في غرف العرض. أيضاً نفترض أن البائع يرغب في معرفة عدد الطرق الممكن بها ترتيب هذه الحراثات من موديل ٥٠ ، وموديل ٦٠ ، وموديل ٧٠ بغرفة العرض. الطريقة الوحيدة لعمل الخيارات الممكنة لحل هذه المشكلة هو وضع قائمة بالتوليفات الممكنة من هذه الحراثات.

يمكن إيجاد ست توليفات، أو تباديل، هي ٥٠ ٦٠ ٧٠، ٥٠ ٧٠ ٦٠، ٦٠ ٥٠ ٧٠، ٦٠ ٧٠ ٥٠، ٧٠ ٥٠ ٦٠، ٧٠ ٦٠ ٥٠. ويمكن استخدام الرسم الشجري لعرض هذه التوليفات بيانياً. الشكل الشجري (الشكل رقم ٣,١) يوضح كل ترتيب والذي يمكن به عرض الحراثات في غرفة العرض وكما نلاحظ أن كل ترتيب يختلف عن الآخر.



الشكل رقم (٣,١). الشكل الشجري للتباديل الممكنة لثلاثة أنواع من الجرارات.

كبديل لذلك، يمكن اعتبار المتاح في كل خطوة لاختيار الجرار الزراعي ومن ثم الوصول لعدد التباديل الممكنة . في الخطوة الأولى هناك ثلاث خيارات ممكنة لاختيار الجرار الزراعي. ولكن باختيار الجرار الأول يتبقى لدينا خيارات للخطوة الثانية. بعد اختيار جرارين يتبقى لدينا اختيار جرار واحد في الخطوة الثالثة. ولذلك فإن لدينا ثلاث طرق لعمل الخطوة الأولى، طريقتين لعمل الخطوة الثانية، طريقة واحدة لعمل الخطوة الثالثة، حاصل ضرب ٣×٢×١ ، أو ست تباديل أو طرق لاختيار الجرارات الزراعية. حيث تم استخدام قاعدة الضرب، وكما هو متوقع فإن العديد من مشكلات العد تتطلب أكثر من ثلاث خطوات. لذلك ، فإن الأمر يتطلب وضع قاعدة الضرب، بشكل عام. حيث يمكن عمل الخطوة الأولى بعدد n_1 طريقة، والخطوة الثانية بعدد n_2 طريقة، والاستمرار حتى r خطوات ، ولذلك فإن إجمالي عدد الطرق يمكن صياغته بإيجاد حاصل ضرب تلك الخطوات كما في المعادلة رقم (3.1).

$$(n_1)\ (n_2)\ (n_3)\ \ldots\ldots\ldots\ (n_r) \qquad (3.1)$$

على سبيل المثال ، قائمة الغداء في مطعم تحتوي على ثلاثة خيارات من المتبلات ، وخمسة خيارات من الأطباق الرئيسية ، وأربعة أنواع من المشروبات وستة أنواع من الحلويات. ولذلك فإن العدد الإجمالي للتشكيلات المختلفة من قائمة الغداء الممكن طلبها هي :

$$(3)\ (5)\ (4)\ (6) = 360$$

ويمكن عادة تحديد عدد التباديل باستخدام أساسيات الضرب ولكن هناك بعض الحالات التي لا يمكن استخدام هذه القاعدة فيها لإيجاد عدد التباديل الممكنة.

وعلى العموم فإنه في حالة شمول جميع عناصر المجموعة في التباديل يمكن حساب عدد التباديل الممكنة باستخدام المضروب $n!$ كما في المعادلة (3.2).

$$_{n}P_{n} = n! \qquad (3.2)$$

ويمكن توضيح ذلك من خلال المثال التالي : إذا كان المطلوب إيجاد عدد الطرق الممكنة لتوزيع عدد أربعة عمال زراعيين على عدد أربع وظائف مختلفة يمكن استخدام قاعدة المضروب ومن ثمَّ تكون القيمة الممكنة هي :

$$_{4}P_{4} = 4!\ or\ (4)\ (3)\ (2)\ (1) = 24$$

إذاً لدينا ٢٤ توليفة لتعيين العمال في الوظائف.

في بعض الحالات الأخرى عندما يكون الهدف تحديد البدائل لمجموعة جزئية r من المجموعة n فإنه يمكن استخدام الصيغة الموضحة في المعادلة رقم (3.3).

$$_{n}P_{r} = n!/(n-r)! \tag{3.3}$$

ويمكن توضيح ذلك من خلال المثال التالي : إذا كان لدى شركة تعمل في الكيمياء الزراعية عدد ٦ مندوبين مبيعات يعملون في قسم المبيعات بالشركة وترغب الشركة في اختيار عدد ٢ منهم لتعيينهم في وظائف خارج المبيعات مع إبقاء الأربعة الآخرين في القسم فإن عدد الخيارات الممكنة لمندوبي المبيعات تكون :

$$_{6}P_{2} = 6!/(6-2)! = \frac{6!}{4!} = \frac{(6)(5)4!}{4!} = 30$$

وفي حالات أخرى قد يكون الهدف هو إيجاد عدد التباديل لنشاطين متنافيين الحدوث ، على سبيل المثال ، عند وقوع أحد الحدثين فإن الآخر لا يقع. في هذه الحالة إذا تمت صياغة النشاط الأول بعدد j طريقة والنشاط الثاني بعدد k طريقة فإنه يمكن عمل أحدهما أو الآخر بعدد k + j طريقه. مثال ذلك ، إذا كان المزارع يرغب في شراء جرار زراعي وأخبره مندوب المبيعات بتوفر نوعين من الجرارات الأول عبارة عن دفع ثنائي ومتوفر بعدد ست سرعات ، وناقلين للحركة ، وثلاثة أنواع من المظلات ، بينما النوع الثاني يمتاز بدفع أمامي ومتوفر بعدد أربع سرعات وثلاثة نواقل للحركة ، ونوعين من المظلات وعليه فإن لدى المستهلك عدداً من الخيارات لاختيار جرار يمكن إيجادها كالتالي :

$$j + k = (6)(2)(3) + (4)(3)(2) = 60$$

التوافيق Combinations

بما أن التباديل تعتمد على تنظيم الترتيبات ، بحيث أن أي تغيير في التنظيم يعطى

ترتيب مختلف ، إلا أن التوافيق لا تعتمد على التنظيم. وعليه فإن عدد التوافيق الممكنة لـ n عناصر في مجموعه معينه يساوي واحد ؛ لأن الترتيب لا يهم ولذلك لدينا طريقة واحده لشمول جميع العناصر في التوفيق. ولكن يمكن اختيار مجموعة جزئية r من المجموعة n ومن ثم الحصول على أكثر من توفيق والتي يمكن تعريفها بالرمز التالي $_nC_r$.

في المثال التالي ، نفترض أن لدينا أربعة عناصر ، $w\,x\,y\,z$ ، ونرغب في إيجاد عدد التباديل والتوافيق لمجموعة من ثلاثة عناصر في الوقت نفسه من هذه المجموعة. يوضح الجدول رقم (٣,١) قائمة بالخيارات الممكنة للتباديل والتوافيق ولذلك يمكن ايجاد أربع من عدد التوافيق لعدد ثلاثة عناصر مختارة من مجموعة مؤلفة من أربعة عناصر ، ويمكن إيجاد عدد ٢٤ من التباديل.

ولذلك فإن كل توفيق يحتوي على عدد ٦ تباديل

$$r! = (3)(2)(1) = 6$$

وعند قسمة العدد الإجمالي لعدد التباديل $n!/(n-r)!$ على عدد التباديل لكل توفيق $r!$ نحصل على عدد التوافيق الممكنة والتي يمكن صياغتها رياضياً في المعادلة التالية (3.4) .

$$_nC_r = n!\,/\,r!\,(n-r)! \qquad (3.4)$$

وكمثال على ذلك ، فلو افترضنا أن مدير المبيعات يرغب في وضع أربعة أنواع مختلفه من الجرارات في غرفة العرض على أرضية المعرض من عدد ستة أنواع موجودة لديه فإن عدد التوافيق الممكن عرضها يمكن حسابها كالتالي :

$$_6C_4 = 6!/(4!)(6-4)! = 15$$

وسـيتم اسـتخدام التوافيـق لاحقـاً في حالـة دراسـة التوزيـع الاحتمـالي ثنـائي الحدين.

الجدول رقم (٣،١). مقارنة عدد التباديل والتوافيق الممكنة للحروف wxyz في مجموعة من ثلاثة عناصر.

التباديل						التوافيق
yxw	*ywx*	*xyw*	*xwy*	*wyx*	*wxy*	*wxy*
zxw	*zwx*	*xzw*	*xwz*	*wzx*	*wxz*	*wxz*
zyw	*zwy*	*yzw*	*ywz*	*wzy*	*wyz*	*wyz*
zyx	*zxy*	*yxz*	*yxz*	*xzy*	*xyz*	*xyz*

الاحتمال Probability

تـستخدم الاحتمـالات عنـد الحاجـة لـصنع قـرار حيـال ظـاهرة ذات نتـائج غـير مؤكـدة. تعـرض نظريـة الاحتمـالات الطريقـة الرياضـية لتقـدير النتـائج غـير المؤكـدة، ونرغـب في الاختيـار بـين بعـض الاسـتنتاجات مـن الاحتمـالات والتفـسير الرياضـي المتحصل عليه من مسلمات الاحتمال.

ولكـن يجـب ذكـر بعـض المفـاهيم قبـل الـدخول في ذلـك ، وأول هـذه المفـاهيم التجربة. حيث تعرّف التجربة بأنها عبارة عن أي عملية يتم إجراؤها بغـرض الملاحظة وهكذا ، فإن رمي زهرة النرد، أو ملاحظة عـدد بـذور الطمـاطم النامية في الأراضـي الطينية أو تسجيل عدد العملاء في السوق المركزي عند نقط المحاسبة عبارة عن تجـارب. والتجربة لها نتائج . العـدد الـذي يظهـر عنـد رمـي زهـرة الـزد، نمـو بـذرة معينـة ، عـدد العمـلاء عنـد نقطـة المحاسـبة في الـسوق المركـزي خـلال الـساعة الأولى مـن يـوم الـسبت عبارة عن نتائج محتملة للتجربة.

الحدث الأولي، أو أي حدث، عبارة عن الاسم الذي يتم إعطـاؤه لكـل نتيجـة للتجربة التي يمكن أن تقع في محاولة مفردة. على سبيل المثال، محاولـة التجربـة المتضمنة

لرمي زهرة النرد تمثل النتائج الممكنة أعداد النقاط من ١ إلى ٦ الموجودة على الأوجه المختلفة وبذلك فإن أي رقم يظهر بعد رمي الحجر عبارة عن حدث ، والأعداد من ١ إلى ٦ عبارة عن جميع الأحداث الممكنة لتجربة رمي زهرة النرد. والأحداث السابقة عبارة عن حوادث متنافية ؛ حيث إن وقوع أحدها ينفي وقوع الأحداث الأخرى. وكمثال آخر في حالة وجود محل تجاري لديه نقطتين للمحاسبة فإن الأحداث متنافية ؛ نظراً لأن المستهلك سيحاسب عند محاسب واحد فقط ، الحوادث لها مشاهدات أو وحدات أولية مرتبطة بها ، وجمع الوحدات الأولية معاً يعطي المجتمع.

نفترض أن لدينا صندوقاً يحتوي على مسامير بقطر $\frac{1}{4}$ بوصة ومقاسات أخرى وتم إجراء تجربة لاختيار مسمار بطريقة عشوائية من الصندوق وملاحظة هل المسمار المختار قطره $\frac{1}{4}$ بوصة أم لا؟ ثم إعادته في الصندوق والنتيجة المحتملة مرتبطة بالوحدات الأولية.

يتم وضع التفسيرات للاحتمالات الموضوعية بطريقتين مختلفتين لمعنى الاحتمالات ثم توظيف إما النظرية وإما الخبرة لتحديد الاحتمالات. الأحداث المتساوية تشير إلى أن جميع نتائج التجربة لها نفس الفرصة في الظهور باحتمالات متساوية وهذه الطريقة تستخدم في الحالة التي لا يوجد هناك سبب لتفضيل نتيجة معينة للتجربة. وعليه فإن احتمال ظهور أحد وجهي العملة (إما الكتابة أو الصورة) يساوي $\frac{1}{2}$ لأن هنالك نتيجتين للتجربة ويظهر وجه واحد عند رمي العملة ويكون للوجهين نفس الفرصة في الظهور.

والاحتمال $\frac{1}{2}$ عبارة عن نسبة عدد الطرق التي يمكن أن يظهر بها وجه معين مقسوماً على إجمالي النتائج الممكنة للتجربة. وتتيح لنا النظرية حساب الاحتمالات دون إجراء عملية الرمي للعملة حيث يمكن تحديد الاحتمالات باستخدام النظرية ،

أوالحكم الحدسي وتسمى الاحتمالات المسبقة. وقبل تطبيق فكرة نتائج ذات احتمالات متساوية يجب الحصول على احتمالات مثل أن يكون لدينا أوراق لعب متساوية أو حجرة نرد سليمة ، وبذلك فإننا لا نحتاج لتجربة في مثل هذه الحالات.

ولكن عند إجراء تجربة في ظروف متماثلة فإن النتائج المتحصل عليها من المحاولات تكون بتكرارات مختلفة. وفي حالات أخرى لا يمكن معرفة الاحتمالات المسبقة ببساطة ولهذا تم ابتكار طريقة للحصول على الاحتمالات تحت هذه الشروط تسمى التكرار النسبي والذي يعبّر عن عدد مرات وقوع حدث مؤكد في n محاولة للتجربة ومن ثم الحصول على تقدير عددي لاحتمال الحدث. ولسوء الحظ فإن التقدير يعتمد على عدد المحاولات في التجربة. ولكن وجد أنه عند إجراء عدد كبير من المحاولات في التجارب التي تعرف فيها الاحتمالات المسبقة فإن التكرار النسبي يقترب من القيمة الأولية. وهذا يعطينا يقين باستخدام التكرار النسبي لقياس الاحتمالات ، ولذلك يمكن تعريف احتمال الحدث A , $P(A)$ بالمعادلة التالية (3.5) :

$$P(A) = \frac{\text{عدد النواتج المحققة للحادث A}}{\text{عدد النواتج الكلية في التجربة}} \tag{3.5}$$

خصائص الاحتمال الأساسية

تتصف الاحتمالات بثلاث خصائص أساسيه هي :

١- إذا كان الحدث A أحد الأحداث المتنافية أو الشاملة لتجربة معينه فإن قيمة الاحتمال $P(A)$ والتي تعبّر عن نسبة عدد مرات وقوع الحدث A إلى عدد مرات إجراء التجربة يجب أن تقع بين الصفر والواحد . أي أن :

$$0 \leq P(A) \leq 1$$

٢- نظراً لأن الحوادث تجميعية (شاملة) فإن أحد الحوادث الأولية يجب أن تقع عند إجراء محاولة ولذلك فإن احتمال وقوع حدث غير الحدث *A* (الحدث المكمل) يساوي $P(A')=1-P(A)$. وعليه فإن مجموع الاحتمالات لجميع الحوادث يساوي واحد.

٣- يعبّر الحدث A' عن مجموعة من الحوادث المركبة متنافية الحدوث غير الحدث A ولذلك تسمى الحوادث المركبة. وعليه فإن احتمال الحدث A' ، $P(A')$ ، عبارة عن مجموع الاحتمالات لجميع الحوادث الأولية غير الحدث A .

وعلى سبيل المثال ، وجد أحد مصانع الآيس كريم أنه في المتوسط ٦٠٪ من عملائه يشترون فانيليا ، ٢٠٪ يشترون شوكولاته ، ١٠٪ يشترون فراوله و١٠ ٪ يشترون باقي النكهات الأخرى. ولذلك فإن احتمال أن يشتري مستهلك تم اختياره عشوائياً فانليا هو $P(V)=0.6$ واحتمال اختيار المستهلك لنكهة أخرى غير الفانليا $P(V')$ عبارة عن مجموع احتمال باقي الحوادث.

$$P(C) + P(S) + P(O) = 0.2 + 0.1 + 0.1 = 0.4$$

وعليه فإن احتمال الحدث المركب ، غير الفانليا $P(V')= 0.4$ أو

$$1-P(v') = 1-0.6 = 0.4$$

وأي حدث يساوي احتماله صفر فإنه يعني أنه لا يتوقع حدوثه بشكل كبير جداً ولكن لا يعني استحالة وقوعه ، وبالمقابل فإن احتمال $P(A)=1$ لا تعني أن الحدث مؤكد وقوعه ولكن هناك احتمال كبير جداً بوقوعه.

وتتميز إجراءات التكرار النسبي للاحتمالات بأربع ميزات أساسية هي :

١- عدد كبير من المحاولات.

٢- تقترب قيمة التكرار النسبي من القيمة الأولية في حالة وجودها.

٣- استخدام المعلومات التجريبية المكتسبة من الخبره.

٤- استخدام التكرار النسبي لتقدير الاحتمال.

وفي بعض الحالات يكون هناك محددات للخاصيتين الأول ؛ نظراً لأنه يجب تحديد الاحتمالات قبل إجراء التجربة وفي هذه الحالة يستخدم التقدير الشخصي بناء على الخبرة ولكن بشكل عام استخدام التكرار النسبي واسع الانتشار.

المجموعات Sets

المجموعة عبارة عن مجموعة من الأشياء المعرّفة الموضوعة سوياً وقد تكون حقيقية أو تخيلية. ولذلك فإن موظفي مصاعد الحبوب يشكلون مجموعة. وبنفس الطريقة فإن المجموعة يمكن أن تتألف من النظريات المتعلقة بالبحث عن أصل العالم، أو مجموعة المزارعين الذين لديهم تأمين على محاصيلهم، أو مجموعة النتائج المتحصل عليها من تجربة معينه. والأمثلة السابقة عبارة عن مجموعات منفصلة والتي تتكون من عدد من العناصر المنتهية.

والنقاش التالي يهتم بالمجموعات المنفصلة والتي يمكن وصفها بطريقتين :

١- تحديد جميع عناصر المجموعة.

٢- تحديد مدى انتماء شيء معين للمجموعة بتطبيق القاعدة.

وعند تحديد عناصر المجموعة تتم كتابتها باستخدام قوسين {} وشمول جميع العناصر فيها، فعلى سبيل المثال يمكن تحديد مجموعة أوراق اللعب كعناصر للمجموعة في الشكل التالي {شرياء ، سبيت ، ديمن ، هاص} = S ومجموعة جميع الأحداث المقترنة بتجربه تسمى المجموعة الشاملة ، S .

العمليات على المجموعة Set Operations

في معظـم الحـالات يكـون الاهتمـام بجـزء (أو مجموعـة جزئيـة) مـن عناصـر المجموعة الكلية. أي حدث عبارة عن مجموعة جزئية من المجموعة الشاملة للتجربة[١] والأحداث يمكن أن تحتوي على مجموعة من العناصر كما في المثال التالي . نفترض أنه تم رمي زهرتي نرد معاً ولذلك فإن هناك ست وثلاثين نتيجة محتملة للتجربة ويكون الناتج عبارة عن عدد النقاط التي ظهرت على وجهي الحجرين في الرمية. فإذا تم تعريف الحدث بأنه عبارة عن مجموعة النتائج التي يكون مجموعها ٧ يمكن إيجاد المجموعة الجزئية التالية :

$$S_1=\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$$

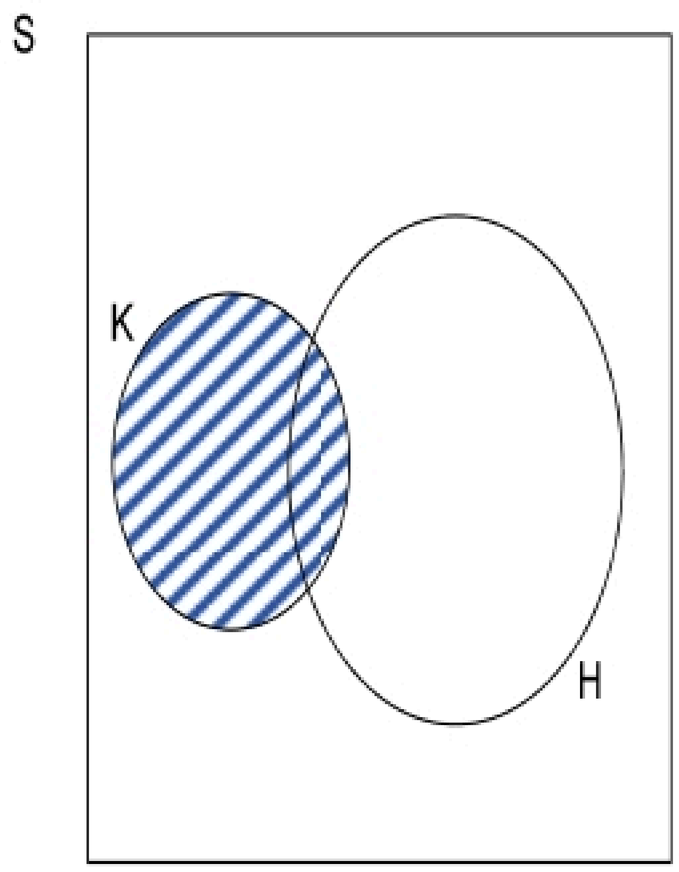
عند وجود طريقه لتميز كل زهره عن الأخرى مثل استخدام حجرين بلونين مختلفين.

وفي بعض الأحيان فإنه يمكن تحديد المجموعات الجزئية بسهولة من المجموعة الشاملة عند استخدام الرسم. ويعتبر شكل فن Venn Diagram أحد الطرق لعرض المجموعة الشاملة بمستطيل والمجموعة الجزئية بدائرة داخل المستطيل. فمثلاً لو اعتبرنا أن المجموعة S هي مجموعة أوراق اللعب (عدد ٥٢) يمكن تعريف مجموعتين جزئيتين منهما المجموعة الجزئية الأولى H وهي تمثل الأوراق التي تحتوي على الأوراق التي على شكل القلب H والمجموعة الجزئية الثانية هي التي تحتوي على الأوراق التي عليها صورة الملك K .

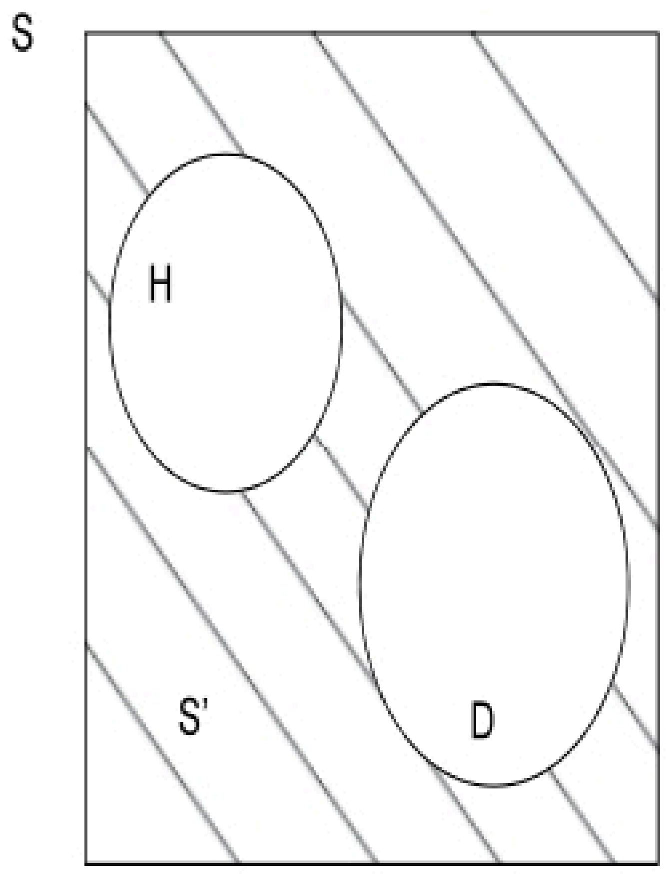
في هذه الحالة يمكن عرض المجموعة الكلية S باستخدام المستطيل كما في الشكل رقم (٣,٢)[٢] والمجموعات الجزئية H أو K يمكن عرضها باستخدام الدوائر داخل المستطيل . المساحات المظللة و/ أو المساحات غير المظللة في الشكل تعبّر عن عناصر

المجموعات الجزئية سواء في المجموعة H أو المجموعة K والتي تمثل اتحاد المجموعتين H و K وتكتب $H \cup K$ وتقرأ H أو K أو H اتحاد K. بينما المساحة الناتجة عن تداخل الدائرتين تعبّر عن العناصر المشتركة في المجموعتين الجزئيتين ، أي التي يكون فيها الملك والقلب معاً. وتعّرف بأنها عبارة عن تقاطع المجموعة H و K وتكتب $H \cap K$ وتقرأ H و K ، أو H تقاطع K . وتسمى المجموعة الجزئية التي تحتوي على عنصر واحد مثل $H \cap K$ بمجموعة الوحدة.

ويمكن أن يتم اختيار مجموعات جزئية من المجموعة الشاملة لا تحتوي على عناصر مشتركة. وعند التمثيل البياني لتلك المجموعات فإنها لا تتقاطع ؛ حيث إن العنصر في أحد المجموعات لا يمكن أن يكون عنصر في المجموعة الجزئية الأخرى ؛ لأنها منفصلة عن بعضها ، وتعبّر المساحة المتبقية من الشكل عن المجموعة الشاملة S وغير المشمولة في أي من المجموعات الجزئية عن المجموعة المكملة للمجموعة S . ويعبر عنها بالرمز S' .كمثال لذلك :



الشكل رقم (٣,٢). الرسم التوضيحي للملوك والقلوب.



الشكل رقم (٣,٣). الرسم التوضيحي للفئات المنفصلة والمكملة.

يوضح الشكل رقم (٣،٣) شكل فن لمجموعة أوراق اللعب S وتم اختيار مجموعتين جزئيتين الأوراق التي تحمل القلب H والأوراق التي تحتوي الديمن D حيث عبر عنها بالدائرتين H ، D. وكما يلاحظ ليس هناك تداخل بين المجموعتين الجزئيتين H ، D ؛ لأنه لا يوجد عنصر مشترك بينهما ولذلك فإن باقي أوراق اللعب التي لا يتضمنها الديمن D أو القلب H ، $S\text{-}(HUD)$ ، تشكل المجموعة المكملة S' والتي تم توضيحها بالشكل المخطط الباقي من المستطيل وتجدر الإشارة بأنه سيتم التطرق للمجموعات وخصائصها، الاتحاد والتقاطع والتكامل لاحقاً.

المتغيرات العشوائية Random Variables

نتذكر بأن المتغير عبارة عن شيء يأخذ مدى معين من القيم . وفي دراسة الاحتمالات تمت الإشارة لفكرة المتغير عندما نهتم بملاحظة القيم عند رمي زهرتي نرد أو الاختيار العشوائي للمزارعين من الإطار. في هذه الحالات، نحن لا نعرف مسبقاً النتيجة للتجربة ولكن لدينا بعض المعلومات عن التكرار النسبي للنتيجة.

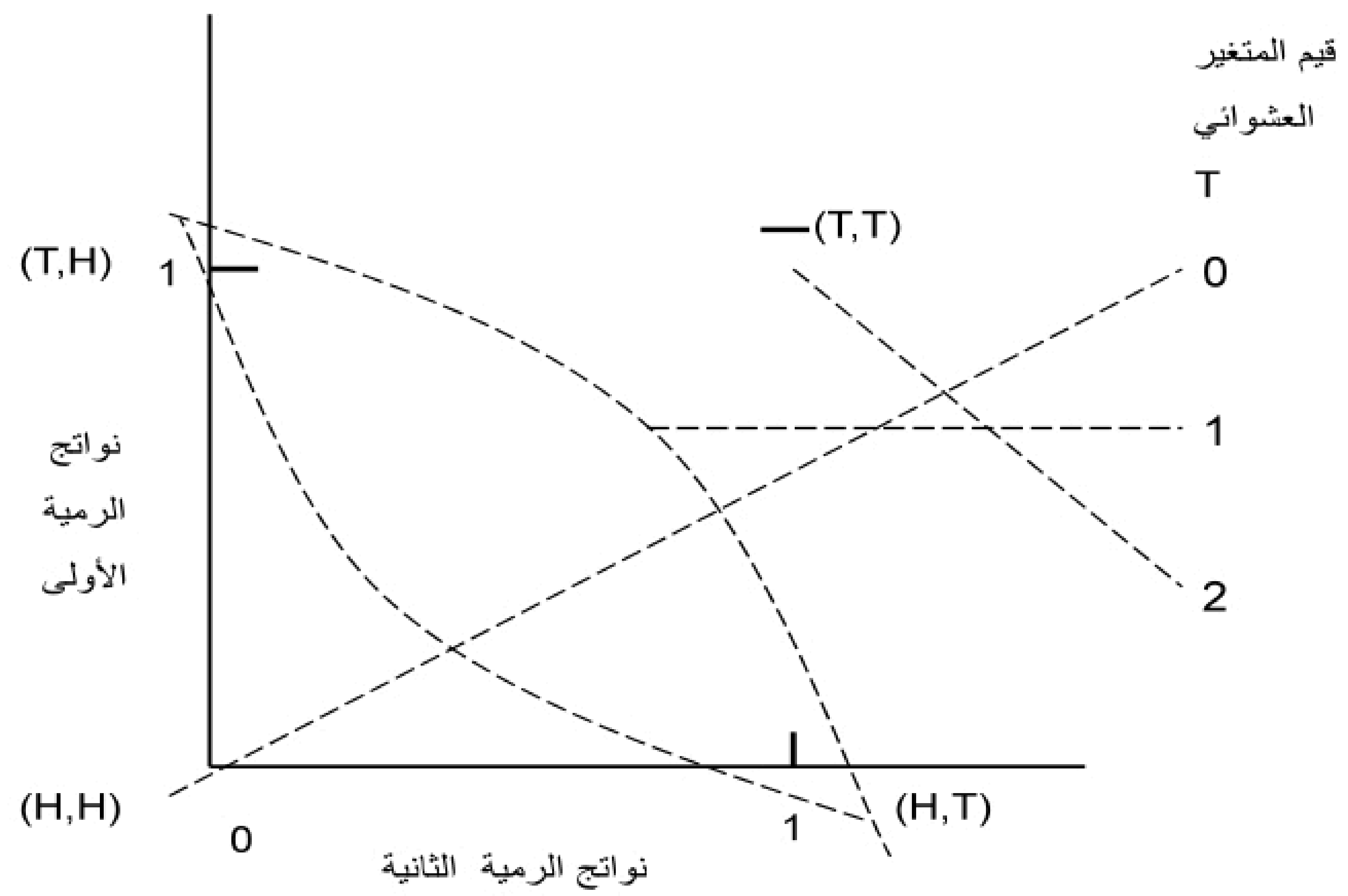
وعندما تكون الحوادث غير متشابهه أو لا يمكن التنبؤ بها فردياً بدرجة من الثقة فإنها توصف بالعشوائية أو الصدفة. وبإعادة المحاولات لتجربة عشوائية نحصل على نتائج أو حوادث. المتغير العشوائي يعبر عن قيمه مؤكده أو خاصية معيّنه لكل حدث فردي موضع الدراسة (ويدعى كذلك متغير الصدفة). وينقسم المتغير العشوائي إلى قسمين، متغير عشوائي متصل ومتغير عشوائي منفصل. وسيتم التطرق للمتغير العشوائي المنفصل أولاً.

باستخدام نظرية المعاينة نستطيع تحديد الأحداث الأولية أو نقاط العينة من أي بيانات تم جمعها. فراغ العينة لأي تجربه عبارة عن جميع نقاط العينة. وكل نقطة مرتبطة

بحدث واحد محتمل في التجربة. وعلى سبيل المثال، بافتراض أنه تم رمي عمله معدنية سليمة مرتين كتجربة عشوائية. في كل رميه يمكن الحصول على نتيجة واحدة من النتيجتين المحتملة إما صورة H أو كتابة T وبذلك فإن فراغ العينة لتلك التجربة هو :

$$S = \{(H,H), (H,T), (T,H)\ (T,T)\}$$

وعليه فإن عدد وجوه الكتابة الممكن الحصول عليها T هي { ٠ ، ١ ، ٢} وهو عبارة عن متغير عشوائي منفصل. ويمكن تمييز عناصر فراغ العينة بالقيم المختلفة للكتابة في المجموعة الجزئية $\{(H,H), (H,T), (T,H)\ (T,T)\}$. وعليه فإن الأرقام المحتملة للكتابة في تجربة رمي العملة مرتين مشمولة في المجموعة الجزئية للأحداث في فراغ العينة. ويمكن توضيح فراغ العينة للتجربة بيانياً، الشكل رقم (٣،٤)، مع عرض قيم المتغير العشوائية في عمود مجاور للرسم. الخطوط في الرسم تشير للدالة أو قاعدة العدّ.



الشكل رقم (٣،٤). التمثيل البياني للفراغ العيني للمتغير العشوائي T عند رمي عملة متزنة مرتين.

وعلى أية حال فإن أكثر الاهتمام يتعلق بمعرفة الدالة التي تربط قيم المتغير العشوائي بالاحتمالات المصاحبة. ويمكن تقييم هذه الدالة بمقارنة عدد المرات للنتائج التي تظهر فيها الكتابة $T=1$ أي ظهور الصورة مرة واحدة يساوي $\frac{2}{4}$ أو $\frac{1}{2}$ ، عدم ظهور الصورة، أو $T = 0$ عبارة عن $1/4$ ، وعدد مرات ظهور الصورة مرتين هي مرة واحدة أو $1/4$ ويمكن بناء على ذلك كتابة صيغة تلك الدالة رياضياً بالمعادلة رقم (3.6) والتي تعبّر عن التوزيع الاحتمالي ثنائي الحدين والتي سيتم التطرق له بالتفصيل لاحقاً.

$$P(r) = {}_nC_r\, P^r\, q^{n-r} \tag{3.6}$$

ويمكن التعبير عن العلاقة بين المجموعة الجزئية لفراغ العينة لأي تجربة وقيم المتغير العشوائي باختيار المعادلة المناسبة. وفي الحقيقة فإن محاولة إيجاد أو تحديد الدالة المناسبة لفراغ العينة مهمة صعبة في علم الإحصاء.

قواعد الاحتمال Rules of Probability

القواعد الثلاثة التي تمت الإشارة إليها سابقاً تساعدنا في تفسير القياسات للاحتمالات ولكنها ليست شاملة لدراسة جميع مشكلات الاحتمالات التي يمكن أن تواجهنا في القطاع الزراعي. لذلك سيتم اعتبار بعض قواعد الاحتمالات الإضافية. القواعد الثلاث السابقة يمكن الإشارة إليها بهدف التذكير وهي :

١- $0 \leq P(A) \leq 1$

٢- $P(A_1) + P(A_2) + \ldots + P(A_n) = 1$

٣- $P(A') = 1 - P(A)$

القاعدة الرابعة هي :

$$P(B) = P(A_1) + P(A_2)$$

وهذه قاعدة خاصة للجمع والتي تستخدم فقط في حالة الحوادث المتنافية فإذا كان لدينا عدد k أحداث متنافية يمكن تعميم هذه القاعدة كما في المعادلة رقم (3.7) :

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k) \qquad (3.7)$$

على سبيل المثال، نفترض حظيرة بها عدد ٤٠٠ عجل ، بحيث إن ١٢٠ فقط أنجبت بواسطة ثور A، و٨٠ بواسطة الثور B و ٩٦ بواسطة الثور C و ١٠٤ بواسطة الثور D . وتم اختيار أحد العجول عشوائياً ، فإن احتمال أن أبوه الثور A أو B أو D هو :

$$P(A \cup B \cup D) = P(A) + P(B) + P(D) = \frac{120}{400} + \frac{80}{400} + \frac{104}{400} = \frac{304}{400} = 0.76$$

وباختيار العجل عشوائياً فإن جميع العجول في الحظيرة لها فرص متساوية في الاختيار باحتمال $\frac{1}{400}$. فإذا كان الهدف اختيار عجل تم إنجابه بواسطة الثور A يمكن حساب الاحتمال كالتالي :

$$P(A) = 120\left(\frac{1}{400}\right) = \frac{120}{400} = 0.30$$

وإذا كانت المجموعات الجزئية المزمع دراستها فيها عناصر مشتركة كما في الشكل رقم (٣,٢) فإنه في هذه الحالة الأحداث غير متنافية. فإذا كانت المجموعات متقاطعة

وغير منفصلة نطبق القاعدة العامة للجمع الموضحة في المعادلة رقم (3.8) :

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2) \qquad (3.8)$$

وبالعودة للشكل رقم (٣،٢) فإنه من السهولة ملاحظة أنه يجب أن نطرح التقاطع ؛ لأننا أضفناه مرتين. والقاعدة العامة للجمع يمكن تطبيقها على الاحتمالات سواء أكانت متنافية أم غير متنافية. فعندما تكون الأحداث متنافية فإن المجموعات تكون منفصلة ولذلك فإن التقاطع يساوي صفراً، وعليه فإن القاعدة العامة للجمع في هذه الحالة تتحول إلى القاعدة الخاصة للجمع. وقواعد الجمع مفيدة في إيجاد احتمالات الحدث المركب . ولكن بعض مسائل الاحتمالات يتطلب الأمر فيها معرفة الاحتمال المشترك لمجموعة أحداث تقع معاً فإذا كانت الأحداث مستقلة، فإن احتمالها المشترك يمكن إيجاده باستخدام القاعدة الخاصة بالضرب ، والموضحة في المعادلة رقم (3.9) لعدد K من الأحداث المستقلة A_1 , A_2 , . . . , A_k

$$P(A_1 \cap A_2 \cap . . . \cap A_K) = P(A_1).P(A_2). ..... .P(A_K) \qquad (3.9)$$

وكمثال لذلك ، إذا كانت سجلات الجامعة المحلية تشير إلى أن ثلث طلاب تخصصات الزراعة شقر ونصفهم طلبة دراسات عليا مستجدين وثلاثة أرباعهم حضروا المحاضرات أكثر من سنه. ولذلك فإن احتمال اختيار طالب تخصص من الكلية يكون أشقر وطالب دراسات عليا مستجد له في الجامعة أكثر من سنه بافتراض أنها أحداث مستقلة يكون :

$$(\frac{1}{3})(\frac{1}{2})(\frac{3}{4}) = 3/24 = \frac{1}{8}$$

الأحداث المستقلة Independent Events

يكون الحدثان A ، B ، مستقلان إذا تحقق الشرط في المعادلة رقم (3.10) :

$$P(A|B) = P(A) \tag{3.10}$$

وتوضح أن فرصة وقوع الحدث A ، بمعرفة الحدث B ، مساوية لفرصة وقوع الحدث A بدون معرفة الحدث B . أو أن معرفتنا بالحدث B لا يغير احتمال وقوع الحدث A ، ولذلك فإن الحدث A مستقل إحصائياً عن الحدث B . ومن جهة أخرى ، إذا كان لدينا حدثين A ، B وكانت العلاقة بينهما كما هو موضح في المعادلة رقم (3.11) فإن الحدث A يعتمد على الحدث C .

$$P(A|C) \neq P(A) \tag{3.11}$$

حيث إن معرفة الحدث C تغيّر من احتمال الحدث A ويسمى $P(A|C)$ احتمال الحدث A المشروط بوقوع الحدث C .

الاحتمال الشرطي Conditional Probability

في بعض الأحيان نشاهد بعض المسائل التي يتم التعامل فيها بنسبة من فراغ العينة بدلاً من كامل فراغ العينة. وبذلك فإن احتمال الحدث يعتمد على الجزء من فراغ العينة الذي تم اختياره. على سبيل المثال، في حالة اختيار جميع الموظفين في مؤسسة أعمال زراعية كبيره الذين لديهم درجة جامعية يختلف عن الموظفين التجاريين ذوي الدرجات الجامعية الذين يتم اختيارهم من تلك المؤسسة. حيث إن الموظفين الإداريين عبارة عن مجتمع جزئي معرّف بشروط خاصة يتم اختيارهم من المجتمع . ويمكن تعريف الاحتمالات المرتبطة بتلك الحوادث في المجتمع الجزئي بالحوادث الشرطية. فمثلاً

نفترض أن تاجر السيارات المحلي لديه نوعان من السيارات سيدان وسيارات رياضية ومتوفرة كلا النوعين بناقل حركة عادي و أوتوماتيكي كما في الجدول رقم (٣,٢).

الجدول رقم (٣,٢). أعدد السيارات السيدان والسيارات الرياضية (SUVs) مصنفة حسب نوع ناقل الحركة.

نوع ناقل الحركة	سيارات سيدان Sedans	سيارات رياضية(SUVs)	الإجمالي
عادي	٢	١	٣
أوتوماتيكي	٤	٣	٧
الإجمالي	٦	٤	١٠

في حالة توفر جميع السيارات لدى التاجر فإن فراغ العينة S عبارة عن ١٠ خيارات يمكن اختيار أيٍّ منها بشكل عشوائي. فلو تم تعريف الحدث A بأنه اختيار سيارة بناقل حركة أوتوماتيكي يكون احتمال الحدث $P(A)=0.7$ ويكون الاحتمال للحدث B والذي هو تفضيل عشوائي للسيارة الرياضية يساوي $P(B)=0.4$. ولكن بافتراض أن المستهلك اختار شراء سيارة رياضية، في هذه الحالة فإن فراغ العينة (s) انخفض من $S = 10$ إلى فراغ جزئي يساوي $B = 4$.

وبذلك فإن اختيار الحدث A والذي هو عبارة عن سيارة بناقل حركة أوتوماتيكي المشروط بأن تكون سيارة رياضية يمكن حسابه كالتالي :

$$\mathrm{P}(A\,|B)= \frac{3}{4} = 0.75$$

ويمكن تأكيد صحة هذه النتيجة باعتبار أن الحدث المشروط $\mathrm{P}(\mathrm{A}|B)$ مساوٍ لـ $P(A \cap B)$ ، 3/10 ، والتي تعبّر عن نسبة السيارات ذات ناقل الحركة

الأوتوماتيكي ومن نوع السيارات الرياضية مقسوماً على $P(B)$ ، 4/10 ، والـذي يعبّر عن نسبة جميع السيارات الرياضية كما في المعادلة رقم (3.12) :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/10}{4/10} = \frac{3}{4} = 0.75 \qquad (3.12)$$

مع ملاحظة أن $P(B) \neq 0$.

القواعد العامة للضرب General Rule of Multiplication

إذا كـان لـدينا حـدثان مـستقلان A ، B يمكـن التعـبير عـن احتمـال وقوعهمـا أو احتمالها المشترك بأحد العلاقات الموضحة في المعادلة رقم (3.13) :

$$P(A \cap B) = P(B) \;.\; P(A \backslash B) \qquad (3.13)$$

أو

$$P(A \cap B) = P(A) \;.\; P(B \backslash A)$$

ويمكن اعتبار المثال التالي والذي يفترض أن أحد العاملين بمتجـر لبيـع الحفـارات الكهربائية ويرغب في الاختيـار مـن عـدد سـتة حفـارات متـشابهه وهـذه الحفـارات فيهـا أربعه صالحة واثنان غير صالحة وغير معروف لديه أي منها صالح أو غير صالح.

فـإذا أختـار العامـل حفـارين عـشوائياً فمـا احتمـال أن كـلا الحفـارين صـالحة ؟ وحيث إن السحب من تلك العينة يكون بدون إرجاع للحفار المختار ، فـإن احتمـال أن ذلك الحفار يعمل $P(w_1) = 4/6$ ويكـون الاحتمـال الـشرطي أن الحفـار المختـار ثانيـاً يعمــل إذا إن الحفــار الأول الــذي تم اختيــاره صــالح يكــون : $P(W_2 \backslash W_1) = 3/5$ وبذلك فإن الاحتمال المشترك للحفارين أنها صالحة يكون :

$$P(W_1 \cap W_2) = P(W_1) \, . \, P(W_2 \mid W_1) = (4/6)(3/5) = \frac{12}{30} = 2/5$$

التوقع الرياضي Mathematical Expectation

مفهوم التوقع الرياضي أحد المفاهيم التي سيتم التطرق لها في الفصول الأخيرة. ويمكن ربطه بفكرة الاحتمالات التي سبق مناقشتها . القيمة المتوقعة ، أو التوقع الرياضي للمتغير هي عبارة عن المتوسط الحسابي للمتغير مرجحاً بالاحتمالات. حيث إن الأوزان عبارة عن الاحتمالات المرتبطة بكل قيمه للمتغير العشوائي. فمثلاً نتذكر من العرض السابق لتجربة رمي عملة معدنية غير متحيزة مرتين عدد مرات ظهور الكتابة T ، كمتغير عشوائي تكون النتائج الممكنة هي ٠ ، ١ ، ٢ باحتمال $\frac{1}{4}$ و $\frac{2}{4}$ و $\frac{1}{4}$ على التوالي وبذلك تكون القيمة المتوقعة للمتغير العشوائي عبارة عن مجموع حاصل ضرب كل قيمة للمتغير العشوائي باحتمالها كما في المعادلة رقم (3.14).

$$E(T) = \sum T_i \, . \, P(T_i) \tag{3.14}$$

وبتطبيق ذلك نحصل على القيمة المتوقعة لعدد مرات ظهور الكتابة :

$$E(T) = (0)(\frac{1}{4}) + (1)(\frac{2}{4}) + (2)(\frac{1}{4})$$

$$E(T) = 0 + \frac{1}{2} + \frac{1}{2} = 1$$

وعليه فإن القيمة المتوقعة تساوي واحد ، أي أننا نتوقع في المتوسط ظهور الكتابة مرة واحدة في حالة رمي عمله معدنية غير متحيزة مرتين.

وقد تكون القيمة المتوقعة مساوية لأحد قيم المتغير العشوائي أو غير مساوية لها. حيث إن تفسيرها يعبّر عن المتوسط عند إجراء التجربة لعدد كبير من المرات، أو المتوسط المرجح بالاحتمالات لقيم المتغير العشوائي.

وبصفه عامة فإن فكرة التوقع الرياضي قد طوّرت واستخدمت في ألعاب الحظ ولكنها أصبحت الآن من الأدوات المتعارف عليها والتي تستخدم في كثير من المجالات لتقييم الخيارات المختلفة.

ملاحظة ختامية Endnote

١- لاحظ أن هذا غير مماثل للحدث الأولي، كما عرف سابقاً في مفهوم تحليل الاحتمالات. لسوء الحظ، كل من نظرية المجموعات ونظرية الاحتمال تحتوي على أحداث.

تمارين Exercises

١- نفترض أن لدينا ثلاث مجموعات A و B و C تحتوي على العناصر التالية:

B={4,5,6,7,8,9},   C={4,6,8,10}   A={3,4,5},

المطلوب:

أ) حدد المجموعات الجزئية من المجموعة A. ثم أوجد المجموعة المكملة للمجموعة الجزئية $\{4,5\}$.

ب) أوجد $A\cap B$ و $A\cup B$. ثم استخدم رسم شكل فين لعرض هذه المجموعات.

ج) أوجد $A\cup B\cup C$ و $A\cap B\cap C$ و $(A\cup B)\cap C$ ثم أعرض هذه المجموعات باستخدام شكل فن.

٢- افترض أن لديك فراغ العينة التالية :

$$S = \{2 , 4 , 6 , 8 , 10\}$$

والمجموعات الجزئية التالية :

$$A = \{2 , 4 , 6\}$$
$$B = \{6 , 8\}$$
$$C = \{4 , 6 , 8 , 10\}$$

المطلوب إيجاد المجموعات التالية :

أ) $A \cap B$ (ب) $A \cup C$ (ج) $(A \cup C)' \cap B$ (د) $A \cap C'$

٣- إذا رمي حجر نرد ٣٠٠ مرة، كم عدد المرات التي تتوقع ظهور الرقم ٣ فيها ؟ ما هو احتمال الحصول على عدد أكبر من ٣ ؟

٤- إذا تم رمي حجرين نرد معاً فأوجد الاحتمالات التالية :

أ) ظهور العدد ٧.

ب) ظهور عدد زوجي.

ج) ظهور العدد٧ أو عدد زوجي.

د) ظهور العدد ٧ أو العدد ١١.

هـ) ظهور عدد أقل من ٧ .

و) تقاطع العدد أقل من ٧ وعدد زوجي.

٥- افترض وجود صندوق يحتوي على أربع رقائق تم ترتيبها من ١ إلى ٤ المطلوب :

أ) إذا تم اختيار رقيقتان مع الإرجاع، أوجد فراغ العينة.

ب) احتمال أن يكون مجموع الأرقام على العينة المسحوبة في المثال السابق أقل من ٥ .

ج) احتمال أن يكون مجموع الأرقام على العينة المسحوبة في المثال السابق أقل من ٥ إذا علمت أن الرقيقة الأولى هي التي تحمل الرقم ٢.

٦- افترض سحب ورقة لعب بدون إعادة من مجموعة أوراق لعب عددها ٥٢ ورقه مخلوطة بعناية المطلوب:

أ) احتمال أن تحصل على ملك وسبيت في سحبتين متتاليين.

ب) احتمال أن تحصل على خمس سبيت في خمس سحبات متتالية.

ج) في سحبتين متتاليتين ، ما احتمال الحصول على العدد ١٠ في السحبة الثانية إذا علمت أن العدد ١٠ أو الجوكر سحبت في السحبة الأولي.

د) في لعبة البوكر، ما احتمال الحصول على ٤ وحدات وبعض الأحداث الأخرى.

٧- أ) كم عدد الطرق الممكن استخدامها لعرض أربعة من العجول في صالة العرض من الأول إلى الرابع.

ب) كم عدد الطرق الممكن استخدامها لعرض ستة من العجول في صالة العرض من الأول إلى الرابع.

ج) كم عدد التوليفات الممكنة التي يمكن بها عرض أربعة عجول مختارة من مجموعه عددها ستة.

٨- موزع تجهيزات زراعيه لديه عدد آلتين لربط البالات الدائرية وأربع الآت لربط البالات المربعة. يرغب في عرض هذه الآلات في صف في المعرض، كم عدد الطرق المختلفة الممكن استخدامها لعرض هذه الآلات.

٩- يتكون نادي الطلاب من عشر طالبات واثنا عشر طالب، يرغب النادي في تشكيل مجموعة مكونه من ستة طلاب يتم اختيارهم عشوائياً . ما احتمال أن يكون أربعة أعضاء من المجموعة المختارة طالبات.

١٠- إذا كان احتمال ولادة الذكر والأنثى متساوي ، ما هو احتمال ولادة عدد اثنين من الذكور في مجموعه من الحيوانات عددها ثمانيه ؟ ما احتمال ولادة ذكرين على الأكثر ؟

١١- ارسم التوزيع الاحتمالات للمتغير العشوائي X ، عدد الصور في حالة رمي عمله متزنة ثلاث رميات، ثم كوّن جدول يحتوي على قيم المتغير العشوائي والاحتمالات المصاحبة لها. احسب القيمة المتوقعة للمتغير العشوائي X . ثم ارسم شجرة القرارات التي توضح جميع الاحتمالات الممكنة.

١٢- قرر نادي طلاب القيام بعملية اليانصيب لجمع النقود. تم طباعة ٥٠٠ تذكرة كتب عليها قيمة ٢ دولار مع احتمال الفوز ببندقية صيد قيمتها ٢٠٠دولار:

أ) ما هو التوزيع الاحتمالي للمتغير العشوائي X ، قيمة الربح لمشتري تذكره تم اختياره عشوائياً، إذا تم بيع جميع التذاكر المصدرة.

ب) احسب القيمة المتوقعة لـ X.

١٣- الجدول التالي يوضح إنتاج إحدى محطات التجارب الزراعية لزراعه الجزر والذي تم حصاده وتدريجه ثم فرزه حسب اللون.

	الدرجـــــــــــة			
اللون	الأولى	الثانية	الثالثة	المجموع
برتقالي	٩٠٠	٤٠٠	٢٠٠	١٥٠٠
أرجوان	٦٠٠	٣٠٠	١٠٠	١٠٠٠
المجموع	١٥٠٠	٧٠٠	٣٠٠	٢٥٠٠

احسب الاحتمالات التالية لمجموعة تم اختيارها عشوائياً :

أ) أن تكون لون برتقالي من الدرجة الثالثة.

ب) أن تكون من الدرجة الثانية.

ج) أن تكون من الدرجة الأولى إذا علمت أن لون الجزر أرجواني.

د) أن تكون من الدرجة الثانية أو لون أرجواني .

هـ) أن تكون لون ارجواني ، إذا علمت أنها من الدرجة الثانية.

١٤- إذا كان لدينا قطعه معدنية بها ثمانية وثلاثون فتحة معدّة لسقوط الكرة فيها وكانت تلك الفتحات مرقمه بـ ، ٠ ، ٠٠ ، وأرقام من ١- ٣٦ بحيث إن لون الفتحات المرقمة بالأصفار خضراء والأرقام الفردية لونها أحمر والأرقام الزوجية لونها أسود وتم رمي الكرة عشوائياً احسب الاحتمالات التالية :

أ) احتمال سقوط الكرة في الفتحات السوداء.

ب) احتمال سقوط الكرة في الفتحة رقم ١٠.

ج) احتمال سقوط الكره في فتحه سوداء أو الفتحة رقم ١٠ .

د) احتمال سقوطها في فتحه حمراء أو خضراء.

هـ) إذا كنت ترمي الكرة لمرات غير منتهية وترغب سقوطها في الفتحات الحمراء كم تتوقع أن يكون الاحتمال الممكن الوصول إليه.

الفصل الرابع

التوزيعات الاحتمالية

Probability Distributions

في بعض الأحيان لا يقتصر الاهتمام على معرفة الاحتمالات المصاحبة لحدث معين ولكن يتعدى ذلك إلى معرفة توزيع تلك الاحتمالات على كامل المدى لجميع الأحداث الممكنة. على سبيل المثال، لو افترضنا زهرة نرد مرقمه من ١ - ٦ وتم تعريف المتغير العشوائي X بأنه الرقم الذي يظهر عند رمي الزهرة يمكن كتابة الاحتمالات المرتبطة بكل قيمه من قيم المتغير العشوائي X كما في المعادلة رقم (4.1) وجميع الاحتمالات متساوية لافتراضنا أن الزهرة سليمة من العيوب :

$$P(X=1) = 1/6, \quad P(X=2) = 1/6, \ldots, P(X=6) = 1/6 \qquad (4.1)$$

وتسمى قيم المتغير العشوائي X والاحتمالات المصاحبة لها بالتوزيع الاحتمالي ويمكن استبدال كتابة كل احتمال مصاحب في حالة تساوي الاحتمالات بالصيغة الرياضية التالية:

$$P(X) = 1/6 \text{ for } (X = 1, 2, \ldots, 6) \qquad (4.2)$$

ويكون مجموع قيم الاحتمالات المصاحبة للمتغير X مساوي للواحد.

وهذه الخاصية تشمل جميع التوزيعات الاحتمالية، وكما هو معلوم فإن الاحتمالات يجب أن تكون موجبه أو غير سالبة.

من جهة أخري فإن هناك علاقة بين التوزيعات الاحتمالية والتوزيعات التكرارية ، حيث إن التوزيعات التكرارية عبارة عن قائمة تشمل جميع النتائج للمتغير التي تم تقسيمها إلى فئات مع التكرارات لكل فئة.

في حين أن التوزيع الاحتمالي يعبر عن جميع النتائج الممكنة للمتغير العشوائي مع الاحتمال المصاحب لكل قيمة بدلاً من التكرار.

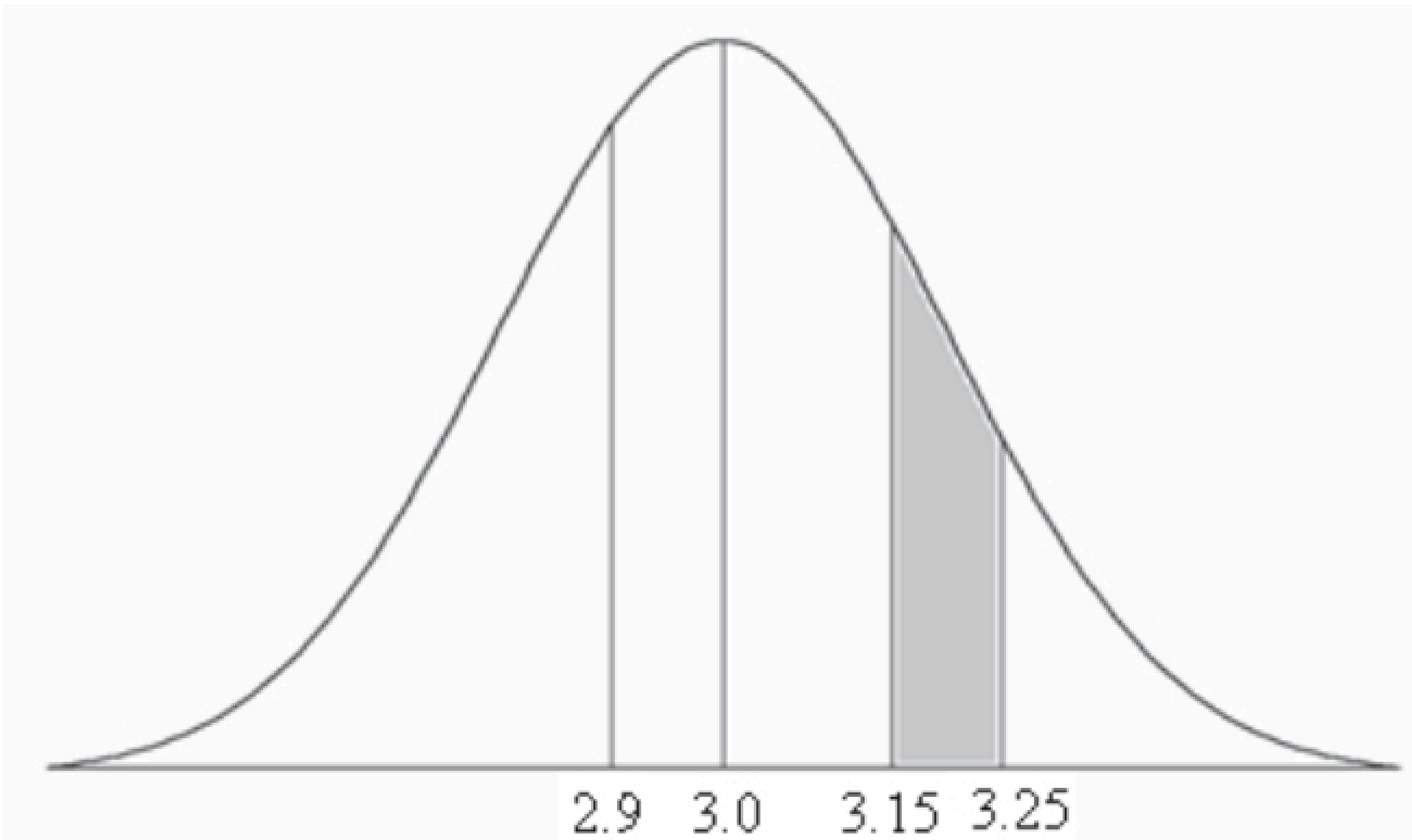
وتهدف العديد من المشكلات الإحصائية إلى اختيار التوزيع الأفضل المناسب للمتغير العشوائي كما تمت الإشارة لذلك سابقاً. وقبل اختيار التوزيعات الاحتمالية المستخدمة غالباً في الإحصاء، سيتم النظر إلى دوال الكثافة الاحتمالية.

دوال الكثافة الاحتمالية Probability Density Functions

دالة الكثافة الاحتمالية، ويرمز لها بـ pdf ، قد تكون منفصلة أو متصلة. وتُعرّف دالة الكثافة الاحتمالية المنفصلة بأنها عبارة عن دالة نقطيه على فراغ عينة منتهي ، وبذلك فإنها تأخذ قيم عددية منتهية فقط.

وتشير الـ pdf للاحتمال المصاحب لكل ناتج ممكن للمتغير العشوائي المنفصل. أما دالة الكثافة الاحتمالية المتصلة فهي مجموعة دوال تعرض التوزيع الاحتمالي للنواتج الممكنة لمدى من القيم للمتغير العشوائي المتصل.

ويوضح الشكل رقم (٤،١) قيمة الاحتمال بيانياً للمتغير العشوائي المتصل X ، الذي يعبّر عن قطر عمود المحرك المنتج من شركة الآلات الزراعية بقطر يتراوح بين٣،١٥ إلى ٣،٢٥ والذي هو عبارة عن المساحة المظللة.



الشكل رقم (٤,١). التوزيع الاحتمالي المتصل.

وتعبر دالة الكثافة الاحتمالية pdf عن مجموعة مستقلة من التوزيع الاحتمالي بحيث تكون قاعدة لتحديد الاحتمالات المصاحبة لأحداث التجربة، بينما التوزيع الاحتمالي عبارة عن عرض منظم أو ترتيب للاحتمالات عند تحديدها. من هذا المنطلق يمكن دراسة بعض التوزيعات الاحتمالية الشائعة الاستخدام.

التوزيع الاحتمالي ثنائي الحدين Binominal Probability Distribution

مجموعة المشكلات الإحصائية التي نصادفها كثيراً تشمل أحداث عشوائية بحيث يكون لها نتيجتين فقط. هذه النتائج تحدث بدون نمط ثابت ولكن احتمال حدوث هذه النتائج يبقى ثابتاً لأي محاولة.

هذا النوع من المسائل يسمى محاولات برنولي. يتكون نظام برنولي من ناتجين أو حدثين فقط، حيث إن النتائج أو الحدث الذي نهتم بدراسته أو نسعى لمعرفته يسمى

نجاح والناتج الآخر يسمى فشل. في بعض الأحيان يكون لدينا معرفه بمستوى من الثقة بمقدار النجاح المتحصل عليه من عدد كبير من المحاولات والذي يدعى المعلمة أو العامل المتغير.

دالة الكثافة الاحتمالية لمحاولات برنولي تسمى دالة الكثافة الاحتمالية ثنائيه الحدين والتي تعطي الاحتمال P لعدد r نجاحات في عدد n محاولة من التجربة. وبالطبع فإن معرفتنا لعدد حالات النجاح r يُمكّننا من معرفه عدد حالات الفشل والتي تساوي $n - r$ ؛ نظراً لأن لدينا ناتجين فقط للتجربة وبذلك يمكن حساب احتمال الفشل والذي يساوي $1- P$ والذي يعبر عنه أحياناً بـq. ويمكن صياغة دالة الكثافة الاحتمالية ثنائية الحدين رياضياً في المعادلة رقم (4.3) :

$$\mathrm{P}\ (\mathrm{r} \setminus \mathrm{np}) = {}_{\mathrm{n}}\mathrm{C}_{\mathrm{r}}\ \mathrm{p}^{\mathrm{r}}(1-\mathrm{p})^{\mathrm{n-r}}\, or\ {}_{\mathrm{n}}\mathrm{C}_{\mathrm{r}}\ P^{r}\ q^{n-r} \qquad (4.3)$$

ويمكن إيضاح ذلك بالمثال التالي : نفرض رمي حجر نرد سليم أربع مرات ونرغب في معرفة احتمال الحصول على رقم ١ ثلاث مرات في هذه الأربع رميات.

ولحل هذه المسألة نحسب أولاً عدد الطرق الممكن الحصول بها على رقم ١ ثلاث مرات في الأربع رميات باستخدام التوافيق حيث

$${}_{4}\mathrm{C}_{3}\ =\ 4!/(4-3)!\ 3!= 4$$

وعليه يمكن حساب احتمال وقوع أي من هذه الطرق باستخدام قاعدة ضرب الاحتمالات. فمثلاً احتمال الحصول على رقم واحد ثلاث مرات بشكل متتالٍ (١) وقيمة أخرى غير الواحد مثلاً (١َ) يمكن حسابها كالتالي :

$$\mathrm{P}(111 1') = (1/6)(1/6)(1/6)(5/6) =\ 5/1296$$

وعليه فإن احتمال الحصول على رقم ١ في المحاولات الثلاث الأولى من أربع محاولات يمكن حسابها كالتالي :

$$(4)(5/1296) = 20/1296 \quad \text{أو} \quad 5/324$$

ويمكن استخدام دالة الكثافة الاحتمالية للتوزيع ثنائي الحدين لحساب ذلك الاحتمال كالتالي :

$$(4)(1/6)^3 (5/6)^1 = 5/324 \quad \text{أو} \quad {}_4C_3\, p^3 q^1 \quad \text{أو} \quad {}_4C_3\, p\, p\, p\, q$$

ويمكن إيضاح ذلك بمثال آخر : وباستخدام المعادلة رقم (4.3) بافتراض أن لدينا عشرة مواليد من العجول ونرغب في معرفة احتمال أنها تحتوي على ذكر واحد. يمكن استخدام دالة الكثافة الاحتمالية لتوزيع ثنائي الحدين حيث $r=1$ ، $n=10$، $P=0.5$ والتعويض في المعادلة (4.3).

$$P(r=1 \mid n=10, p=0.5) = {}_{10}C_1 (0.5)^1 (0.5)^9 = 10!/1!9!(0.5)(0.00195) = 0.0098$$

يمكن تفسير هذه النتيجة بأنها غير متماثلة.

وتجدر الإشارة إلى أن المصطلح ثنائي الحدين أتى من طريقه صياغة دالة الكثافة الاحتمالية. حيث إن قيمتها لأي قيم معطاة لـ r ، n ، p مكافئة للحد العام في مفكوك ثنائي الحدين والذي يعبر عنه بـ $(p+q)^n$.

وتعطي صياغة دالة الكثافة الاحتمالية ثنائية الحدين طريقة حساب الاحتمال لأي عدد محتمل للنجاح في أي عدد معطى من المحاولات، وترتيب تلك الاحتمالات بطريقة معينة يعطي دالة التوزيع الاحتمالي ثنائي الحدين. ولذلك فإن دالة الكثافة الاحتمالية ثنائية الحدين تعرّف كامل التوزيعات، توزيع واحد لكل زوج من قيم n ، P ، معالم تلك التوزيعات.

وحيث إن التوزيعات التي تكون فيها قيمه $P=0.5$ متماثلة ، فإن التوزيعات التي تكون فيها قيم P صغيره تكون متماثلة بصوره معقولة خاصة عندما تكون قيمه n كبيرة.

ومن هذا المنطلق يمكن استخدام المنحنى الطبيعي المتماثل كتقريب لثنائي الحدين والعكس صحيح.

هناك حالات عندما نرغب في إيجاد الاحتمال لعدد (r أو أكثر) من حالات النجاح أو (r أو أقل) في عدد n محاولة. وفي الحقيقة فإننا نهدف إلى إيجاد الاحتمال المصاحب للمساحة تحت أحد أطراف التوزيع ثنائي الحدين. ويمكن الحصول على هذه الاحتمالات بجمع احتمالات ثنائي الحدين لأحداث معينة. وللتجارب المشتملة على عدد كبير من المحاولات ، فإن جمع هذه الاحتمالات يحتاج لوقت ؛ كما أنه مزعج خاصة إذا كانت قيمة r ليست قريبة من الصفر أو قريبه من n . ولهذا السبب تم إيجاد جداول خاصة بتوزيع ثنائي الحدين والتي تعطي مساحات الأطراف للمدى الكبير من قيم n ، P ، r (انظر الملحق جدول رقم (٢) للمثال الجدولي).

الجدول رقم (٤,١). نتائج سبع محاولات لاختيار ذكور وإناث الحيوانات عشوائياً.

المحاولة	الناتج	قيمة X	*تعريف المتغير*
١	F	١	X_1
٢	M	٠	X_2
٣	F	١	X_3
٤	F	١	X_4
٥	M	٠	X_5
٦	F	١	X_6
٧	M	٠	X_7

متوسط وتباين محاولات برنولي

The Mean and Variance of A Bernoulli Process

من السهولة اشتقاق متوسط وتباين توزيع برنولي بطريقه واضحة ؛ نظراً لإمكانية وصف محاولات برنولي بالعديد من الطرق ونفترض أن لدينا تجربة سحب حيوانين مع الإعادة من حظيرة كبيره تحتوي على مجموعة كبيره من الحيوانات عددها n ونصف تلك الحيوانات ذكور M والنصف الآخر إناث F ، ولذلك فإن نسبة الإناث تساوي $P = \frac{F}{n}$. وعليه فإن احتمال النجاح لاختيار حيوان أنثى F تساوي $P = \frac{1}{2}$ ونسبة الفشل (اختيار ذكر M) تساوي $(1 - P) = \frac{1}{2}$. وبدلاً من استخدام دالة التوزيع الاحتمالي ثنائي الحدين في هذه الحالة يمكن استخدام طريقه أخرى. بفرض أننا سحبنا عينة عشوائية من مجموعة حيوانات عددها n في حظيرة وتم تعريف متغير عشوائي منفصل X بحيث تكون قيمته ١ في حاله (سحب أنثى) وقيمته تساوي صفر في حالة الفشل . بهذه الطريقة تم تعريف مجموعة n من المتغيرات العشوائية المستقلة ، $X_1, X_2, \dots, X_n$ بحيث يعبر كل متغير عن محاولة مستقلة ولغرض إيضاح المثال بحيث يكون أكثر وضوح نفترض أننا سحبنا عدد سبعة حيوانات بطريقة عشوائية من الحظيرة وكانت النتيجة كما يلي : $\{F, M, F, F, M, F, M\}$.

وتم عرض هذه البيانات جدولياً في الجدول رقم (٤,١). وبالتالي يمكن حساب القيمة المتوقعة للمتغير X باستخدام المعادلة التالية.

$$\mu = E(X) = \Sigma X \cdot P(X) = (1)(p) + (0)(1-p) = p \qquad (4.4)$$

ويمكن إيجاد التباين للمتغير العشوائي X باستخدام الصيغة الموضحة في المعادلة رقم (4.5) والذي يعبر عنه بـ $Var(X)$ أو σ^2 ويمكن إيجاد التباين للمثال آنفاً كالتالي :

$$\sigma^2 = E(X-\mu)^2 \quad أو \quad E(X-p)^2 = \sum (X-p)^2 \cdot P(X) \qquad (4.5)$$

ويمكن حساب التباين لعدد حالات النجاح r في مجموعة المحاولات n باستخدام المعادلة (4.6)) التي يمكن التعبير عنها كالتالي :

$$\sigma^2 = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p(1-p) \qquad (4.6)$$

وتعبر r عن مجموع المتغيرات العشوائية حيث يمكن إيضاحها في المعادلة التالية:

$$r = X_1 + X_2 + \ldots + X_n \qquad (4.7)$$

ويمكن حساب القيمة المتوقعة لـ r باستخدام المعادلة التالية :

$$E(r) = E(X_1) + E(X_2) + \ldots + E(X_n) = p + p + \ldots + p = np \qquad (4.8)$$

وبطريقة مشابهة يمكن التعبير عن التباين كما في المعادلة رقم (4.9) :

$$\mathrm{Var}(r) = p(1-p) + \ldots + p(1-p) = np(1-p) = npq \qquad (4.9)$$

ويمكن حساب الانحراف المعياري بإيجاد الجذر التربيعي للتباين.

ويمكن إيضاح ما سبق من خلال المثال التالي :

بائع معروضات زراعية لديه مخزن به قطع غيار منها ١٠٪ معيبة فإذا افترضنا أن اختيار القطعة المعيبة هو حالة نجاح وعبّرنا عنه بـ S واختيار قطعة صالحة تعبر عن الفشل وعبّرنا عنه بـ F عليه يمكن حساب القيمة المتوقعة لمرات النجاح لعينة من وحدتين باستخدام المعادلة رقم (4.8) كالتالي :

$$E(r) = np = (2)\,(0.1) = 0.2$$

والتباين يمكن حسابه باستخدام المعادلة رقم (4.9).

$$Var(r) = npq = (2)\ (0.1)\ (0.9) = 0.18$$

وعليه فإن الانحراف المعياري :

$$S = \sqrt{0.18} = 0.42$$

توزيع بواسون الاحتمالي Poisson Probability Distribution

المجموعة الثانية من المسائل الإحصائية يمكن وصفها باحتمالات نجاح صغيرة لأي مجموعة من محاولات عديدة لتجربة معينه. على سبيل المثال هذه المسائل تشمل تطبيقات تأمين المحاصيل، وعدد الكوارث المتعلقة بحالات الطقس في فترة زمنية معينة، وعدد الحراثات القادمة لمخزن الحبوب خلال عملية الحصاد، ومن ثم نوع الدورة أو الصف أو خط الانتظار، والطلب على المواد من مخزن شركة المعدات الزراعية ، عدد العيوب في منتجات مصانع الآلات الزراعية.

ويتميز توزيع بواسون الاحتمالي بملاءمته لعرض هذا النوع من المسائل الإحصائية علاوة على إمكانية استخدام صيغ محدّدة للتوزيع ثنائي الحدين كتوزيع احتمالي لعرض هذه الحالات. وتكون صيغ ثنائي الحدين أكثر ملاءمة عندما يكون عدد محاولات برنولي n كبير واحتمال النجاح p صغير. وفي هذه الحالة فإن الحسابات باستخدام دالة التوزيع الاحتمالي ثنائي الحدين طويلة ومملة ولكنها بسيطة باستخدام دالة التوزيع الاحتمالي لبواسون.

وفي الحقيقة فإن توزيع بواسون لا يمكن وصفه باستخدام أحداث لها ناتجين فقط واحتمال ثابت للنجاح كما في حالة محاولات برنولي. ولكن من خلال شروط معينة يمكن استخدام توزيع بواسون لحلّ مسائل برنولي. على سبيل المثال، هياكل المحراث تصنّع باستمرار بعملية طرق ويكون سمك الجدار للأنبوب أقل من المواصفات المقبولة.

ففي عينة لـ ١٠٠٠ قدم من هذا الأنبوب وجد أن ٢٠ مكاناً منها معيبة (أقل من المواصفات) بصورة عشوائية بمتوسط ٢ لكل ١٠٠ قدم. ففي حالة ثبات احتمال وجود عدد معين من العيوب لكل ١٠٠ قدم يمكن اعتبار مسألة بواسون كحالة برنولي بعرض طول قصير جداً كمحاولة مستقلة بناتج صفر أو أكثر من النجاح.

وكلما كان بالإمكان المحافظة على متوسط np مساوي لـ ٢ معيب لكل ١٠٠ قدم يمكن تجزئة ال١٠٠ قدم إلى ١٠ أجزاء بطول ١٠ قدم أو بطول ١ قدم أو بأي تقسيم مناسب.

وكلما كان عدد القطع كبير فإن احتمال وجود أكثر من عيب في القطعة يكون صغير ويكون الاختلاف بين دالة التوزيع الاحتمالي لثنائي الحدين وبواسون ضئيل.

ولذلك يمكن القول بأنه في حالة مسائل برنولي الاحتمالات مرتبطة بقيمة الاحتمال p للنجاح في عدد n محاولات مستقلة وعدد r من النجاح.

في توزيع بواسون ، يمكن التعبير عن الاحتمالات لعدد معطى من النجاحات l ، لوحدة من المسافة (مثل طول قصير للأنبوب) وعدد حالات النجاح m في قيمة معطاة من المسافة . وتجدر الإشارة بأن الحرف l مماثل للحرف p وكذلك الحرف m مماثل للحرف r . وبذلك فإن العدد المتوقع من النجاحات np لمحاولات برنولي مماثل للـ lm (معادلة رقم 4.10) للعدد المتوقع للنجاحات في قيمة معطاة من المسافة لتوزيع بواسون.

$$\mu = lm \tag{4.10}$$

ويمكن إيضاح ذلك من خلال المثال التالي :

في دراسة لوصول شاحنات الحبوب للرافعات كان احتمال وصول الشاحنات في أي دقيقة مختارة $\boldsymbol{p} = 0.0333$ في حين أن الرقم المتوقع للوصول في نصف ساعة

يمكـن حـسابه كالتـالي $np = (30)(0.0333) = 1$ ويمكـن اختيـار الدقيقـة كوحدة للمسافة l ونصف الساعة كقيمة معطاة للمسافة m وبذلك يمكن حساب الرقم المتوقع u.

$$\mu = (0.0333)(30) = 1$$

ولذلك يمكن حساب احتمالات ثنائي الحدين لعدد r من حالات النجاح في n محاولة باستخدام دالة الكثافة الاحتمالية لثنائي الحدين الموضحة في المعادلة رقم (4.11) التالية :

$$P(r \setminus \text{np}) = {}_nC_r p^r q^{n-r} \qquad (4.11)$$

وكما هو معلوم فإنه عندما تكون قيمه p صغيره فإن التوزيع الاحتمالي سيكون ملتو باتجاه اليمين. وعليه فإنه للقيم الصغيرة للاحتمال p لعدد كبير من المحاولات n مع بقاء ناتج np ثابت فإن نهاية التوزيع ثنائي الحدين هي توزيع بواسون والذي يمكن عرضه رياضياً في المعادل رقم (4.11) التالية :

$$\lim P(r \setminus n, p) = \frac{(np)^r}{r!} e^{-np} \qquad (4.12)$$

حيث e هي الأساس للوغاريثم الطبيعي والتي تساوي ٢,٧١٨٢٨ وبالتعويض عن قيمة np بـ μ يمكن كتابة الصيغة الرياضية لتوزيع بواسون في الشكل التالي :

$$P(r \setminus \mu) = \frac{\mu^r}{r!} e^{-\mu} \qquad (4.13)$$

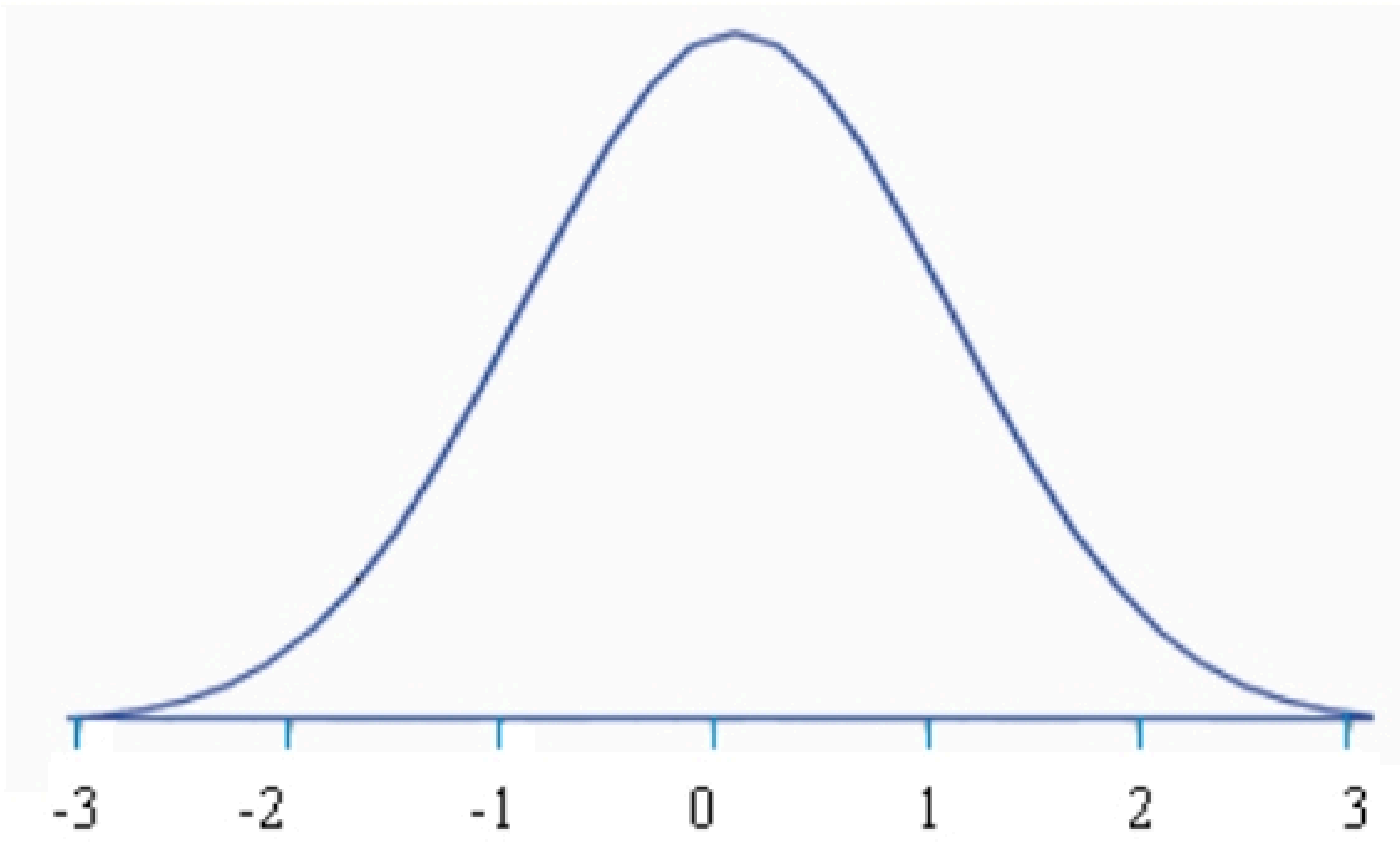
ولذلك فإن هناك معلمة واحدة لتوزيع بواسون والتي يمكن تحديد كامل التوزيع عند معرفتها. حيث إن متوسط التوزيع $E(r) = \mu$ وكذلك

التباين $Var(r) = \mu$. ويمكن إيجاد الانحراف المعياري بحساب الجذر التربيعي للتباين μ . ويمكن الحصول على هذه النتائج مباشرة من التوزيع ثنائي الحدين ؛ حيث إن القيمة المتوقعة $E(r) = np$ والتباين $Var(r) = n\ p\ q$ وتوزيع بواسون الذي يعبّر عن النهاية للتوزيع ثنائي الحدين عندما تقترب n من ما لانهاية و p تقترب من الصفر حيث $np = \mu$ تبقى ثابتة. ولذلك فإن npq تقترب من $np = \mu$ كلما اقتربت قيمة q من الواحد الصحيح.

ويمكن عرض المثال التالي لتوضيح ذلك حيث قام قسم الفحص بمصنع الآلات الزراعية بفحص خرطوم الهيدوليك البالغ طوله ٣٠٠ قدم لمعرفة العيوب وتسجيلها لكل طول من الخرطوم. ومن خبرات سابقه فإن عدد العيوب للطول يتبع توزيع بواسون ، فلو افترضنا اختبار ٢٠ وحدة طول ووجد منها ١٠ تالفة أو معيبة ولذلك فإن تقدير التباين والقيمة المتوقعة $\mu = 10/20 = 0.5$ ويمكن إيجاد الانحراف المعياري بأخذ الجذر التربيعي للتباين والذي يساوي ٠,٧.

التوزيع الاحتمالي الطبيعي The Normal probability Distribution

في حالة المتغيرات العشوائية المنفصلة فإن التوزيع الاحتمالي الأكثر استخداماً هو توزيع ثنائي الحدين في حين أن التوزيع الاحتمالي الطبيعي هو الأكثر استخداماً في حالة المتغيرات العشوائية المتصلة. ويتميز منحنى التوزيع الطبيعي بأنه متماثل ويشبه شكل الجرس ، كما في الشكل رقم (٤,٢). ويقع متوسط التوزيع μ في وسط التوزيع ، كما أنها تتحدد معالمه بالمتوسط μ والانحراف المعياري σ .



الشكل رقم (٤,٢). التوزيع الاحتمالي الطبيعي المعياري.

ونظراً لأن هذا التوزيع نظري أو توزيع حقيقي مرتبط بجميع قيم المجتمع ، فإننا نستخدم الحروف الإغريقية الخاصة بالمجتمع μ ، σ لعرض قيم المتوسط والانحراف المعياري بدلاً من استخدام $\bar{X}$ ، S والتي سوف نستخدمها لاحقاً عندما نتطرق للمتوسط و الانحراف المعياري للعينات المسحوبة من التوزيع.ويتميز التوزيع الطبيعي ببعض الخصائص يمكن إيجازها فيما يلي :

١- المساحة تحت التوزيع الواقعة بين $\mu-\sigma$ و $\mu+\sigma$ تمثل تقريباً ٦٨ ٪ من المساحة الإجمالية الواقعة تحت المنحنى.

٢- المساحة تحت التوزيع الواقعة بين $\mu-2\sigma$ و $\mu-2\sigma$ تمثل تقريباً ٩٥٪ من المساحة الإجمالية الواقعة تحت المنحنى.

٣- المساحة تحت التوزيع الواقعة بين $\mu-3\sigma$ و $\mu+3\sigma$ تمثل تقريباً ٩٩,٧ ٪ من المساحة الإجمالية الواقعة تحت المنحنى.

وعليه فإن مدى المتغير العشوائي X نظرياً يقع بين سالب مالا نهاية وموجب ما لا نهاية تحت المنحنى الطبيعي إلا أنه تقريباً معظم الاحتمالات تتواجد في المدى بين

المتوسط مضاف له ثلاثة انحرافات معيارية والمتوسط مطروحاً منه ثلاثة انحرافات معيارية.

ويتحدد موقع المنحنى وشكله الطبيعي بقيم المتوسط والانحراف المعياري، حيث تحدد قيمة المتوسط مركز التوزيع على خط الأعداد الحقيقية بينما تحدد قيم الانحراف المعياري انتشار المنحنى حول المتوسط. وحيث إن جميع المنحنيات الطبيعية الخاصة بالتوزيع الاحتمالي النظري تتميز بأن إجمالي المساحة تحت المنحنى تساوي الواحد ولذلك فإن زيادة الانحراف المعياري تؤدي لانخفاض ارتفاع المنحنى بحيث يكون مفرطحاً وممتداً أكثر على خط الأعداد. ونظراً لأن شكل المنحنى يتحدد كلياً بواسطة الانحراف المعياري فإنه يمكن معايرة كل المنحنيات الطبيعية باستخدام انحراف معياري واحد بإجراء تعديل بسيط في المتغيرات.وعليه فإن أبسط منحنى يمكن استخدامه للمعايرة بين المنحنيات الطبيعية هو المنحنى الذي متوسطه صفر وانحرافه المعياري يساوي الواحد الصحيح (الشكل رقم ٤,٢) وعليه فعند الحاجة لحساب الاحتمالات لأي منحنى طبيعي يتم أولاً تحويل ذلك المنحى إلى المنحنى الطبيعي القياسي. ويمكن التعبير عن دالة الكثافة الاحتمالية للمنحنى الطبيعي باستخدام المعادلة التالية:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{4.14}$$

وتتسم هذه المعادلة بصعوبة تقييمها ولذلك يستخدم الجدول رقم ٧ بالملحق للحصول على الاحتمالات للتوزيع الطبيعي القياسي. واستخدام هذا الجدول ممكن نظراً لأن أي نقطة على المحور السيني X للتوزيع الطبيعي القياسي يقابلها نقطة على المحور السيني X لأي توزيع طبيعي ويمكن تحديد قيمتها بمعرفة كم انحراف معياري تبعد عن المتوسط.

وعلى سبيل المثال يمكن إيضاح ذلك بافتراض أن لدينا منحنيين تختلف فقط في الانحراف المعياري بحيث إن متوسط المنحنى الأول يساوي صفر وانحراف معياري يساوي الواحد والمنحنى الثاني لمتوسط صفر وانحراف معياري يساوي ثلاثة. وبذلك يمكن تحويل المنحنى الثاني إلى الأول بتغير منحنى X بقسمة كل قيمه على ٣ (قيمه الانحراف المعياري) وعليه فإن قيمة $X = 6$ على المنحنى الطبيعي يعادلها $X = 2$ على المنحنى الطبيعي القياسي وهكذا بالنسبة لباقي القيم.

وبصفة عامه، لأي نقطه x على المنحنى الطبيعي ذي المتوسط μ وانحراف معياري σ هناك قيمه مقابله Z على المنحنى الطبيعي القياسي، وعليه فإن النقطة x تقع على بعد Z انحراف معياري على يمين المتوسط أو $X = \mu + z\sigma$ والتي يمكن صياغتها كمعادلة رياضيه كالتالي:

$$z = (X - \mu) / \sigma \tag{4.15}$$

هذه الصيغة لـ Z تعبّر عن الطريقة المستخدمة لمعايرة القيم للمتغير x والتي تم استخدامها سابقاً. وباستخدام هذه الصيغة الرياضية يمكن تحويل جميع القيم لأي منحنى طبيعي إلى قيم مناظره لها في المنحنى الطبيعي القياسي ومن ثم حساب الاحتمالات تحت أي منحنى طبيعي باستخدام الجدول ٧ بالملحق. فمثلاً بافتراض أن x تنتمي لمنحنى طبيعي بمتوسط ٢٣٠ وانحراف معياري ٢٠ ($x \sim N(230,20)$ ونرغب في حساب احتمال أن قيمه x أكبر من أو تساوي ٢٨٠ ؟

في هذه الحالة يتم إيجاد القيمة المناظرة لقيمة x على المنحنى الطبيعي القياسي Z باستخدام المعادلة رقم (4.15) .

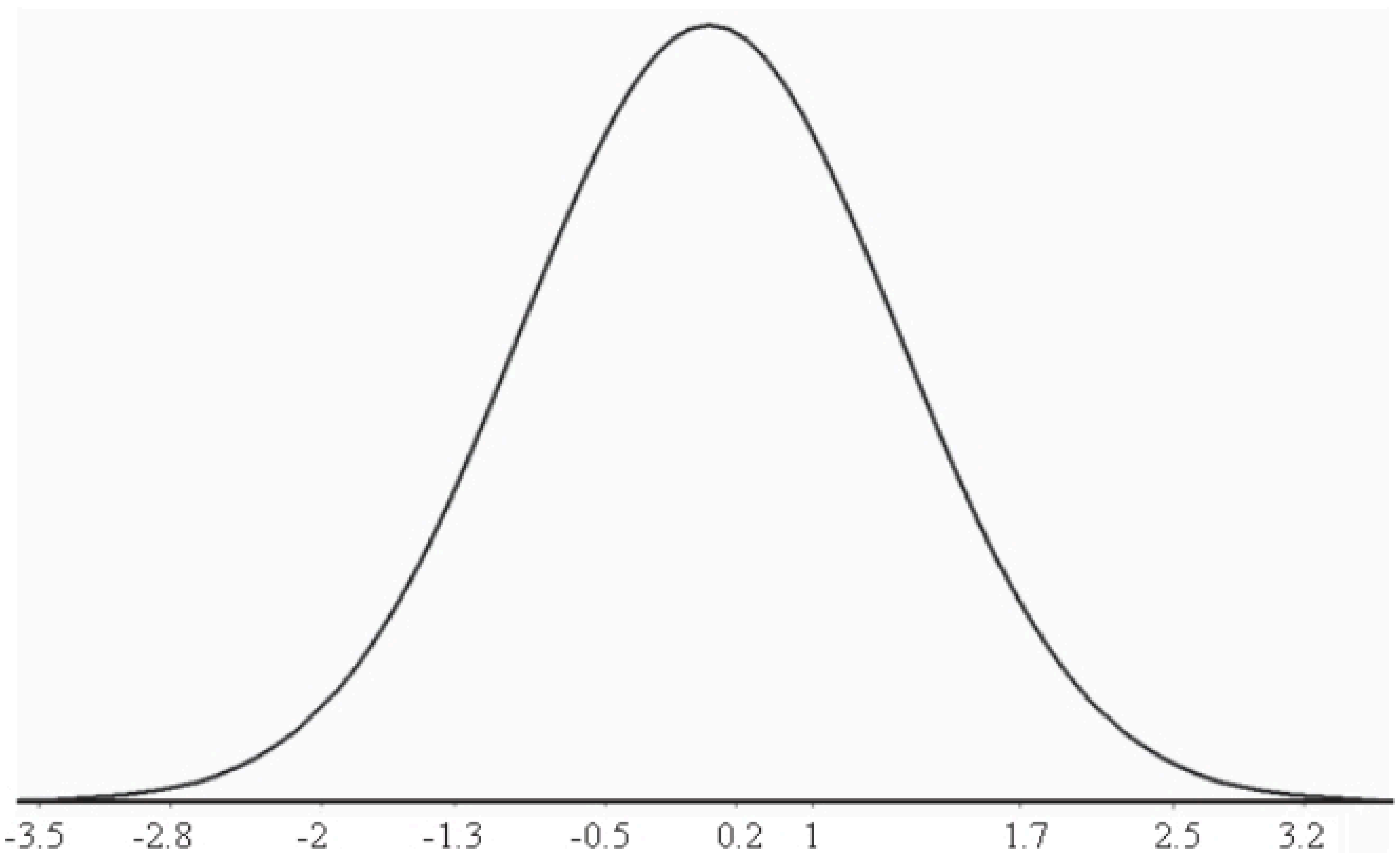
$$z = \frac{280-230}{20} = \frac{50}{20} = 2.5$$

وباستخدام جدول رقم ٧ في الملحق لإيجاد الاحتمال بين قيمة المتوسط وقيمة Z نجد أن الاحتمال يساوي ٠,٤٩٣٨ وعليه فإن إيجاد احتمال أن قيمة X أكبر أو تساوي ٢٨٠ يمكن حسابها بطرح قيمة الاحتمال السابق من ٠,٥ والتي تساوي ٠,٠٠٦٢.

والآن يمكن النظر لمثال آخر ما احتمال أن تكون قيمه X أقل من ٢٧٠ ؟

بنفس الطريقة السابقة يمكن حساب قيمة Z $z = \frac{270-230}{20} = \frac{40}{20} = 2$

من الجدول رقم ٧ نجد أن الاحتمال بين المتوسط وقيمة Z تساوي ٠,٤٧٧٢ والاحتمال يسار المتوسط يساوي ٠,٥ وعليه فإن احتمال أن قيمة أقل من ٢٧٠ يساوي $0.50 + 0.4772 = 0.9772$.



الشكل رقم (٤,٣). التوزيع الاحتمالي الطبيعي المعياري

ويجب ملاحظة أن هذه الاحتمالات عبارة عن المساحات تحت المنحنى وجمع أو طرح تلك الاحتمالات مكافئ لجمع أو طرح المساحات (الشكل رقم ٤,٣) حيث

إن إجمالي المساحة تحت المنحنى تساوي الواحد الصحيح إلا أن إيجاد الاحتمال لقيمة محددة لـ X غير معرّف ؛ لأنه ليس هناك مساحة يمكن إيجادها لنقطه محددة بل يجب أن تكون بين نقطتين.

ودائماً تصاغ مسائل الاحتمالات لمدى من القيم للمتغير العشوائي X عند استخدام التوزيعات المستمرة مثل التوزيع الطبيعي. وفي المثال السابق ليس هناك فرق عند القول أوجد احتمال أن قيمة X أقل من ٢٧٠ ؛ أو القول أوجد احتمال أن قيمة X أقل من أو تساوي ٢٧٠ لأن الإجابة واحده.

التقريب الطبيعي لاحتمالات ثنائي الحدين

Normal Approximation to Binomial Probabilities

عندما يكون لدينا تجربة مشابهة لمحاولات برنولي ونرغب في حساب الاحتمال لـ r من حالات النجاح في عدد n من المحاولات بحيث إن n كبيرة ، يمكن في بعض الأحيان استخدام التوزيع الطبيعي كتقريب لتوزيع ثنائي الحدين. فمثلاً إذا تم اختيار عينة n=20 عشوائياً من عمليات تصنيعية باحتمال يساوي p = 0.4 ونرغب في معرفة احتمال الحصول على خمسة عيوب ، أي r = 5 وباستخدام دالة الكثافة الاحتمالية للتوزيع ثنائي الحدين يمكن الحصول على التالي :

$$P(r=5,\ n=20,\quad p=0.4) = {}_{20}C_5 (0.4)^5 (0.6)^{15} = 0.0746$$

ويمكن حساب التقريب الطبيعي لاحتمال ثنائي الحدين بحساب المتوسط E(r) والانحراف المعياري ثم إيجاد المساحة التقريبية للمنحنى الطبيعي كالتالي :

$$\mathrm{E}(r) = np = 20(0.4) = 8$$

$$\sigma_r = \sqrt{npq} = \sqrt{(20)(0.4)(0.6)} = 2.19.$$

ولإيجاد الاحتمال لقيمة r نستخدم طريقه تقريبية، كما ناقشنا سابقاً أن المنحنى الطبيعي عبارة عن منحنى مستمر، ويصعب إيجاد الاحتمال لنقطة محددّه مثل r = 5 . في هذه الحالة يمكن التغلّب على هذه المشكلة برسم الشكل البياني للتوزيع ثنائي الحدين حيث نجد أن r = 5 هي عبارة عن مركز الفئة التي حدّها الأعلى ٥,٥ وحدّها الأدنى ٤,٥. وباستخدام هذه النقاط كقيم لـ r يمكن إيجاد المساحة للمنحنى الطبيعي والتي تعطي تقريب للمساحة المناظرة وبالتالي الاحتمال الذي نرغب في إيجاده.

وعليه يمكن معايرة القيم بحساب قيمة Z ثم إيجاد الاحتمالات المناظرة باستخدام الجدول رقم ٧ بالملحق. وأخيراً يتم طرح الاحتمالات المتحصل عليها لإيجاد الاحتمال المناظر للمساحة كالتالي.

$$z_1 = \frac{4.5-8}{2.19} = \frac{-3.5}{2.19} = -1.60 \qquad z_2 = \frac{5.5-8}{2.19} = \frac{-2.5}{2.19} = -1.14$$

باستخدام الجدول في الملحق يمكن إيجاد الاحتمالات:

$$z_1 = -1.6 \qquad p = 0.4452$$
$$z_2 = -1.14 \qquad p = 0.3729$$

وعليه فإن الاحتمال:

$$p(4.5 \le r \le 5.5) = 0.4452 - 0.3729 = 0.0723$$

وهذا الاحتمال قريب جداً من الاحتمال الحقيقي لثنائي الحدين. وعليه فإننا حصلنا على إجابة مقاربه باستخدام التقريب الطبيعي. والجدير بالذكر أن طريقة الحساب بصفة عامة أسهل في حالة استخدام التقريب الطبيعي بدلاً من استخدام دالة الكثافة الاحتمالية لثنائي الحدين وخاصة عند الحاجة للاحتمالات التجميعية.

ودالة الكثافة الاحتمالية للتوزيع الطبيعي هي عبارة عن نهاية التوزيع ثنائي الحدين عندما تصبح n كبيرة جداً وعليه فإن التقريب الطبيعي يكون مكافئاً لثنائي

الحدين عندما تكون n كبيرة. وكقاعدة عامة يمكن استخدام التقريب الطبيعي في أي وقت ما دام المتوسط أكبر من ٥ عندما يكون الاحتمال أقل من ٠,٥ و/أو $nq > 5$ والاحتمال أكبرمن٠,٥ وتكون نتيجة التقريب الطبيعي ضعيفة في أطراف التوزيع ثنائي الحدين في حين أنها تعطي نتائج جيدّه كلما كانت حول المتوسط.

تمارين Exercises

١- أعطت شركة العروض التجارية المحلية للزراعة عرض للعملاء بخصم ١ ٪ للعميل الذي يدفع كامل القيمة المستحقة خلال عشرة أيام. في الماضي ٢٠٪ من العملاء دفعوا خلال عشرة أيام وفي الشهر الحالي أصدرت سبع فواتير المطلوب حساب الاحتمالات التالية.

أ) احتمال حصول عميلين على الخصم.

ب) حصول عملين على الأكثر على الخصم.

ج) عدم حصول أي عميل على الخصم.

د) جميع العملاء السبعة حصلوا على الخصم.

هـ) حصول خمسة عملاء أو أكثر على الخصم.

٢- أوجد المتوسط والانحراف المعياري للتوزيع الاحتمالي ثنائي الحدين في السؤال رقم (١).

٣- مسئول الشراء في أحد الأسواق الكبيرة يعرف من خبرته في الماضي أن ٢ ٪ من الطماطم الوارد من المكسيك غير صالح ويجب أن يعدم، قام المسئول باختبار شحنة بعد وصولها بأخذ عينة عشوائية لعدد أربع من الطماطم. المطلوب إيجاد الاحتمالات التالية :

أ) لا يوجد طماطم غير صالح.

ب) عدد ٢ من الطماطم غير صالح.

ج) عدد ٢ من الطماطم غير صالح على الأقل.

د) على الأكثر واحد من الطماطم غير صالح.

٤- من خبرة الطبيب البيطري السابقة فإن ٣٠ ٪ من العجول المصابة بالأمراض يمكن شفاؤها. تم اختبار عدد ستة عجول في مشروع محلي مصابة بالمرض. المطلوب حساب الاحتمالات التالية :

أ) شفاء جميع العجول الستة.

ب) شفاء عجلين على الأكثر.

ج) عدم شفاء أي من العجول.

د) ارسم التوزيع الاحتمالي باستخدام الشكل التخطيطي وحدّد هل هو متماثل أم ملتو ؟ احسب المتوسط والانحراف المعياري.

٥- يتوزع عدد الحوادث الشهرية في شركات تعبئة اللحوم تبعاً لتوزيع بواسون إذا كان متوسط التوزيع ٠,٥ خلال الشهر الأخير. المطلوب حساب الاحتمالات التالية :

أ) عدم وجود حوادث.

ب) حصول حادث واحد.

ج) حصول أكثر من حادث – (استخدام قاعدة المتممة).

٦- تتوزع المكالمات الواردة لشركة زراعية تبعاً لتوزيع بواسون بمتوسط ثلاث مكالمات في الساعة. ما احتمال أن يصل للشركة خلال الساعة القادمة المكالمات التالية :

أ) على الأقل مكالمتين ؟

ب) مكالمة واحدة ؟

ج) أن لا يصل للشركة أي مكالمات؟

٧- يجرى قسم المبيعات بشركة لبيع بذور القطن ٥٠٠ مكالمة يومياً خلال موسم العمل . فإذا كان احتمال عقد صفقة بيع خلال المكالمة يساوي ٠,٠٢ استخدم تقريب توزيع بواسون للتوزيع ثنائي الحدين لإيجاد احتمال الحصول خلال ال ٥٠٠ مكالمة على :

أ) عشر صفقات بيع فقط.

ب) أكثر من خمسة عشر صفقه بيع.

٨- للتوزيع الطبيعي القياسي أوجد الاحتمالات التالية عندما تكون قيمة z :

أ) أقل من ١ .

ب) بين ١ ، ١,٥ .

ج) أكبر من ٢,١٧.

د) بين صفر و - ٠,٨٣ .

هـ) أقل من - ١,٦٦ .

ز) بين - ١,٢ و ٠,٣٤ .

٩- باستخدام جدول التوزيع الطبيعي القياسي أوجد قيمة الاحتمالات التالية :

أ) $p(z < -z_0) = 0.05$

ب) $p(z > z_0) = 0.1628$

ج) $p(-z_0 < z < z_0) = 0.7540$

د) $p(z > z_0) = 0.0485$

١٠- يتوزع وزن الفواكه المعبأة من قبل شركة الفواكه طبيعياً بمتوسط $u = 10$ أوقية وانحراف معياري $\sigma = 0.4$ أوقيه. إذا تم اختيار أحد العبوات عشوائياً ما احتمال أن يكون وزنها :

أ) أقل من عشر أوقيات.

ب) أقل من ١٠,٧ أوقيه.

ج) بين ٩,٦ و ١٠,٧ أوقيه.

١١- يتوزع وزن العجول حديثة الولادة طبيعياً بمتوسط ٤ أرطال وانحراف معياري ٠,٥ رطل. إذا تم اختيار أحد العجول من مجموعة كبيره من العجول حديثة الولادة فما احتمال أن يكون وزنه :

أ) أكثر من ٣,٥ رطل.

ب) بين ٤,٥ و ٥ رطل.

ج) بين ٣,٥ و ٤,٥ رطل.

١٢- إذا كان عدد أزهار الأقحوان المزروعة في أصيص مقاس ٦ بوصه تتوزع طبيعياً بمتوسط ١٦ وانحراف معياري ٢ ما احتمال أن أحد الأصص المختارة من بيت محمي كبير مملؤ بالأصص المزروعة يحتوي على :

أ) أقل من ١٤ زهرة.

ب) أقل من ٢٠ زهرة.

ج) من ١٢ إلى ٢٠ زهرة.

د) ٩٠ ٪ من الأصيص سوف تحتوي على أكثر من .... زهرة.

الفصل الخامس

المعاينة وتوزيعات المعاينة

Sampling and Sampling Distributions

في علم الإحصاء، يتم عموماً عمل الاستدلال عن متغيرات المجتمع موضع الدراسة بناء على تحليل بيانات العينة. من هنا، يتضح أهمية العينات في علم الإحصاء. وبالطبع فإن العينة يجب أن تكون ممثلة للمجتمع إذا تم الاستدلال بناء على إحصاءات تم حسابها من عينة تم اختيارها بدقة.

وهناك نوعان من أساليب المعاينة يمكن استخدامها لاختيار العينات، العينات المبنية على الاحتمال والعينات غير الاحتمالية والتي سيتم التطرق لها بالتفصيل.

العينات المبنية على الاحتمال (العينات الاحتمالية)

Probability-Based Samples

تكون العينة المختارة مبنية على الاحتمال إذا كان أسلوب المعاينة المستخدم يعطي فرصة معلومة لكل مفردة من مفردات المجتمع بالاختيار في العينة. وباستخدام هذا الأسلوب في المعاينة فإنه يمكننا استخدام التعبيرات الاحتمالية في تحليل بيانات العينة. وعند اختيار العينة نفسها فإنه يوجد العديد من الطرق تعطي كل طريقه نوع مختلف من العينات ويعتمد اختيار طريقه المعاينة على مجموعة من العوامل تتمثل فيما يلي:

١- مدى تجانس المجتمع

نظراً لاشتمال المعاينة على أخطاء، فإن طريقة اختيار العينة تعتمد على خصائص المتغير المراد دراسته في المجتمع. فإذا كان المتغير أو الصفة توجد تقريباً في جميع مفردات المجتمع فإن لدينا مجتمعاً متجانساً ويمكن استخدام أسلوب معاينه بسيط مثل العينة العشوائية أو المنتظمة. من جهة أخرى، إذا كانت الصفة متركزة في مستويات مختلفة من خلال مجموعات مختلفة من المجتمع فإنه يجب تحديد تلك المجموعات ثم أخذ عينة من كل مجموعة على حدة. وهذه الحالة تعني استخدام أساليب للمعاينة أكثر تعقيداً مثل العينات الطبقية أو العنقودية والتي تساعد في قياس الاختلافات في المجتمع التي يمكن أن تتسبب في إعطاء قيم مختلفة للمتغير موضع الدراسة.

٢- درجة الدقة المطلوبة

بافتراض تساوي العوامل الأخرى، نرغب في اختيار أسلوب المعاينة الذي يعطي أقل تباين ممكن لإحصاءة العينة لحجم العينة المختارة. من ناحية أخرى، يجب أن نزيد حجم العينة لتقليل خطأ المعاينة والذي يعبّر عن الفرق بين قيمة الإحصاءة أو متغير العينة وقيمة معلمة المجتمع أو متغير المجتمع. وتزيد قيمة هذا الفرق؛ نتيجة للصدفة أو التغيرات العشوائية في اختيار الوحدات الأولية للعينة. وبناء على ذلك، فإن كبر تباين إحصاءة العينة تعني كبر خطأ المعاينة ومن ثمَّ انخفاض الثقة في العينة المختارة.

٣- تكاليف المعاينة

على الرغم من إمكانية زيادة درجة الثقة في العينة بزيادة حجمها إلا أن التكاليف المصاحبة لذلك سوف تزيد بزيادة الحجم. حيث ترتبط تكاليف المعاينة مباشرة بحجم العينة بينما الموثوقية عادة لا ترتبط بشكل مباشر.

وفي الغالب فإن حجم العينة يجب أن يزيد للحصول على زيادة قليله في الموثوقية وبذلك نستطيع تحقيق الكفاءة والذي يعني أن أسلوب المعاينة أكثر كفاءة

مقارنة بأسلوب آخر إذا تم الحصول على نفس درجه الموثوقية بتكاليف أقل. وعليه فإن أسلوب العينة العنقودية عادة أكثر كفاءة من أسلوب العينة العشوائية البسيطة عندما يكون المجتمع المراد أخذ عينة منه منتشر في مساحة واسعة مما يعني تحمل تكاليف سفر عند استخدام أسلوب المعاينة العشوائية البسيطة.

وتصميم العينة يعني تصميم العينة الاحتمالية وبذلك سوف يتم التطرق لأنواع العينات الأكثر استخداماً وتقييم كل منها على حدة بناء على معياري الكفاءة والموثوقية.

المعاينة العشوائية البسطة Simple Random Sampling

في العينة العشوائية البسيطة تتميز جميع عناصر مجتمع المعاينة بتساوي الفرصة في الاختيار. وكذلك في حالة اختيار عينه عشوائياً حجمها n فإن كل عينة محتمله بذلك الحجم لها أيضاً نفس الفرصة في الاختيار.

وتعتبر العشوائية العامل الأساسي المحدّد في حالة العينة العشوائية البسيطة. ولكن ليس بالسهولة تحقيق شرط العشوائية دائماً فعلى سبيل المثال عند الرغبة في استبيان آراء الطلاب في جامعة كبيره حول الزراعة لا نستطيع اختيار عينة عشوائية بسهولة عن طريق مقابلة الطلاب عند باب اتحاد الطلاب حتى نحصل على مجموعة مشاهدات كافية. حيث إن معظم الطلاب ربما لن يذهب لذلك المكان إطلاقاً ولذلك فإن احتمال مقابلتهم سيكون مساوياً للصفر. وعليه فإن العينة المتحصل عليها بهذه الطريقة ليست عينة عشوائية.

وفي الغالب يتم اختيار طريقه المزج الكامل Thorough mixing وذلك لتحقيق العشوائية في العينة، وذلك بترقيم جميع الوحدات في المجتمع أو تعريفها وبذلك لا

يكون هناك فرق بين الوحدة التي تم اختيارها والوحدات الأخرى. حيث يتم وضع الأرقام في سلة كبيره ثم خلطها معاً، يلي ذلك تحديد حجم العينة ثم اختيار الوحدة الأولى ثم الثانية وهكذا حتى نحصل على العدد المطلوب. أما في حالة الاختيار بالإرجاع فإنه يتم خلط الأرقام معاً بعد كل عمليه إرجاع قبل الاختيار حتى نحافظ على نفس الاحتمالية لجميع الوحدات.

تجدر الإشارة إلى أن السحب بدون إرجاع يؤدي لزيادة احتمال الظهور للوحدات المتبقية بعد السحب؛ نظراً لانخفاض العدد الإجمالي للوحدات بعد عملية السحب. ولكن جميع الوحدات المتبقية لها نفس الفرصة في الاختيار مما يحافظ على بقاء احتمالية الحصول على عينه متساوية حجمها n لأي مزيج يتم اختياره من الوحدات والتي يمكن سحبها من مجتمع الدراسة.

ويمكن تبسيط عمليه اختيار العينة باستخدام جدول الأرقام العشوائية والموضح في الملحق (الجدول رقم ١). يحتوي جدول الأرقام العشوائية على أرقام صحيحة موجبة من رقم صفر إلى رقم ٩ ويمكن تركيبها باستخدام طريقة المزج الكامل. حيث يتم إعداد عشرة كروت مرقمه من الصفر إلى ٩ في سلة معينة وسحب واحدة وتسجيلها كأول رقم في الخانة الأولى في جدول الأرقام العشوائية ثم إعادة الكرت ثم السحب مرة أخرى وتسجيل الرقم في الخانة الثانية في الجدول وهكذا حتى نكمل الجدول. وبالطبع فإن الطريقة الأكثر كفاءة هي استخدام برامج الحاسب الآلي لإنشاء الأرقام العشوائية.

باستخدام جدول الأرقام العشوائية نستطيع توليف أي رقمين للحصول على أرقام عشوائية من ٠٠ إلى ٩٩ وثلاثة أرقام للحصول على أرقام من ٠٠٠ إلى ٩٩٩ وهكذا. وبصفة عامة فإننا نختار عدد من الخانات المناسبة للعينة من المجتمع الذي

نرغب دراسته. فعلى سبيل المثال، إذا كان لدينا مجتمع يحتوي على ١٠٠٠٠ وحدة ونرغب في اختيار عينه حجمها ١٠٠، فإنه يجب أولاً وضع قائمة لوحدات المجتمع وترقيمها من ٠٠٠٠ إلى ٩٩٩٩. بعد ذلك نختار مكان للبداية عشوائياً في الجدول ثم نقرأ أفقياً أو للأسفل أو على القطر ونختار أرقام بالتتابع حتى نحصل على١٠٠. في حالة تكرار القيم يتم حذفها والاستمرار في الاختيار حتى نحصل على العدد المطلوب. وأخيراً نختار من مجتمع الدراسة الأرقام التي تم الحصول عليها من الجدول. ويمكن استخدام العينة العشوائية البسيطة إذا كان من ا لسهولة تحديد وحدات المعاينة وكان المجتمع موضع الدراسة صغيراً ومتجانساً. أما في حالة المجتمعات الكبيرة فإن هذه الطريقة مكلفة ومستهلكة للوقت لأنه يجب ترقيم جميع الوحدات. أيضاً إذا كانت الوحدات القريبة من بعضها أكثر تجانساً مقارنة بالوحدات البعيدة كما في استبيانات دخل الأسرة فإن طريقه العينة العشوائية البسيطة قد تعطي عينة غير ممثلة للمجتمع.

المعاينة المنتظمة Systematic Sampling

يتم استخدام العينة المنتظمة بكثرة بدلاً من العينة العشوائية البسيطة إذا كان لدينا طريقه للوصول إلى قائمه بمفردات المجتمع. بهذه الطريقة يمكننا الحصول على عينة بأخذ كل K وحدة في المجتمع، حيث تشير إلى رقم صحيح يساوي تقريباً نسبه المعاينة $\frac{N}{n}$. وبالتالي إذا كان لدينا مجتمع حجمه ١٠,٠٠٠ والعينة n= 500 فإن قيمه k تساوي $20 = \frac{10000}{500}$ وعليه يمكننا اختيار العينة بأخذ كل وحدة عشرين في المجتمع. ولكي يكون لكل وحدة في المجتمع نفس الفرصة في الاختيار يجب اختيار نقطه البداية عشوائياً.

ومن ثم فإن المثال السابق سيتم اختيار رقم عشوائياً بين ١ و ٢٠. فإذا كان على

سبيل المثال الرقم المختار هو ٨ فإن ا لعينه ستحتوي على الأرقام ٨ ، ٢٨ ، ٤٨ ، إلى آخره. حتى نختار جميع الـ ٥٠٠ قيمه. وجدير بالذكر أن استخدام المعاينة المنتظمة ينجح في حالة وجود قائمة بمفردات المجتمع متاحة ، مثل دليل العضوية ، قائمة بعدد المزارعين المقترضين من بنك معين ، إلى آخره. وأحياناً تستخدم العينات المنتظمة في التحكم الإحصائي بالجودة ، وخاصة عندما نختار العينة من جدول الإنتاج الحالي في الفترات المعتادة.

باستخدام طريقة العينة العشوائية المنتظمة ، نحصل على عينة أكثر تمثيلاً مقارنة بطريقه المعاينة العشوائية البسيطة إذا كانت الوحدات في المجتمع المتقاربة أكثر تشابه من الوحدات المتباعدة مثل بيانات الدخل.

من جهة أخرى إذا كانت البيانات تحتوي على فترات أو دورات مخفية بنفس المدى ، كما في فترات المعاينة فإن طريقه المعاينة المنتظمة قد لا تكون الاختيار المناسب. فعلى سبيل المثال ، لا نستخدم طريقة المعاينة المنتظمة لبحث مبيعات محلات المواد الغذائية اليومية وخاصة إذا كنا نأخذ العينة من المبيعات كل سبعة أيام لأن مبيعات المواد الغذائية لا تتوزع على أيام الأسبوع. ولذلك ، لو تم اختيار يوم الثلاثاء سنفقد المتسوقين خلال إجازة نهاية الأسبوع الذين يستفيدون من العروض أو الإجازة من العمل وبالتالي يكون تقديرنا غير صحيح.

المعاينة الطبقية Stratified Sampling

إذا كان لدينا معرفة مسبقة بالمجتمع ويلزمنا تقسيم المجتمع إلى طبقات مختلفة فإن طريقه المعاينة الطبقية أكثر كفاءة من طريقة المعاينة العشوائية البسيطة. ويتم اختيار الطبقات بحيث تحتوي كل طبقة على وحدات لها خصائص متجانسة ومتشابهة أكثر

من المجتمع ككل. بعد ذلك يتم أخذ عينه من كل طبقة باستخدام طريقه العينة العشوائية البسيطة وتكون العينة الإجمالية للمجتمع عبارة عن العينات التي تم أخذها من الطبقات المختلفة. وهناك العديد من الطرق الممكن استخدامها لاختيار عدد الوحدات المختارة في كل عينة فرعية. إحدى هذه الطرق هو استخدام العينة الطبقية المتناسبة بحيث يكون توزيع العينة على كل طبقة متناسب مع عدد وحدات المعاينة منها إلى حجم المجتمع الكلي أي أن عدد الوحدات في العينة المختارة من الطبقة يكون مساوٍ لنفس نسبة عدد وحدات هذه الطبقة في المجتمع. فعلى سبيل المثال، يمكننا أخذ عينة لإنتاجية القطن باستخدام أنواع التربة كطبقات.

الطريقة الأخرى تتم حسب التجانس داخل الطبقات بحيث يتم التوزيع بناء على الانحراف المعياري للطبقة فكلما كانت الطبقة أكثر تجانساً كان الانحراف المعياري أقل ولذلك قل عدد الوحدات المختارة في العينة من تلك الطبقة. وهذا يجعل التناسب عينة طبقية. ومن الأمثلة على ذلك عند الرغبة في أخذ عينة للتكاليف باستخدام حجم المزرعة كطبقات. في هذه الحالة يتم اختيار نسبة صغيره من المزارع الصغيرة مقارنة بالمزارع الكبيرة. وباستخدام الطبقات فإنه يتم تعيين فئات بحيث تكون وحدات المعاينة في كل فئة متماثلة ولكن الفئات مختلفة. ويتم التحكم بنسبة العينة من كل طبقة ولا ندع ذلك للصدفة ولذلك نحقق عينة ممثلة. وإذا كان التباين للخصائص المشاهدة لكل طبقه أصغر مقارنة بكامل المجتمع، كما هو في العادة صحيح، فإن درجه الموثوقيه لحجم عينه محدد سوف تكون عاليه أو الكفاءة لدرجة معينه من الموثوقيه سوف تكون عاليه بسبب اختيار طريقه المعاينة الطبقية. ولذلك فإن طريقة الطبقات مناسبة للتعامل مع المجتمعات غير المتجانسة. في بعض الحالات، يجب أن تقوم بتعيين طبقات متعددة أو طبقات فرعيه وذلك للحصول على أقصى فائدة من الطبقات. فمثلاً يتم تقسيم إنتاجيه

القطن إلى طبقات وفقاً لنوعية الأرض هل هي جافة أم مروية ثم يلي ذلك تقسيم فئات التربة وفقاً لتلك الطبقات.

المعاينة العنقودية Cluster Sampling

المعاينة العنقودية معاكسه تماماً للمعاينة الطبقية. ففي المعاينة العنقودية يتم أولاً اختيار مجموعات من الوحدات الفردية تسمى عناقيد من المجتمع عشوائياً ثم بعد ذلك نختار جميع العينات أو العينات الفرعية من الوحدات وذلك من كل عنقود للحصول على العينة الكاملة. ولتحقيق نتائج جيده فإن الاختلاف بين العناقيد يجب أن يكون صغيراً بقدر المستطاع والاختلاف بين الوحدات داخل العنقود يجب أن يكون كبير. الهدف من ذلك هو تمييز العنقود بصفات مشابهه للمجتمع بحيث يكون أي عنقود عبارة عن عينه ممثلة. عملياً، هذه الشروط ربما يكون من الصعوبة تحقيقها ولكن يجب أخذها في الاعتبار عند تصميم المعاينة العنقودية.

العناقيد تسمى أحياناً وحدات المعاينة الأولية. فإذا كانت جميع الوحدات في العناقيد المختارة مشمولة في العينة، فإن لدينا معاينة من مرحلة واحدة. أما إذا أخذنا عينة فرعيه من وحدات العناقيد فإن لدينا معاينة من مرحلتين. وكذلك هناك معاينة متعددة المراحل. فمثلاً لإجراء دراسة أو مسح لمعرفة وجهة نظر طلاب الكليات الأمريكية تجاه الزراعة، يمكن اختيار الجامعات كعناقيد أو وحدات المعاينة الأولية ثم اختيار الكليات داخل تلك الجامعات كمرحلة ثانية ثم اختيار الطلاب داخل تلك الكليات كمرحلة ثالثة. والميزة الرئيسة للمعاينة العنقودية هو انخفاض تكاليفها مع درجه معينه من الموثوقية. ولذلك، عندما تكون العينة من مجتمع موّزع جغرافياً، كما في المثال السابق حول رأي الطلاب، فإنه يتم أخذ المناطق الجغرافية كوحدات معاينة أولية. وجدير بالذكر أن التباين في العينة العنقودية قد يكون أكبر مقارنة بطرق المعاينة

الأخرى التي تمت مناقشتها، ولكن انخفاض التكاليف يتيح لنا أخذ عينه بحجم أكبر وتكاليف أقل. وعليه فإن كبر حجم العينة سيؤدي لتقليل التباين بدرجه تجعل طريقه المعاينة العنقودية أكثر كفاءة لكل دولار يتم إنفاقه.

المعاينة التتابعية (التسلسلية) Sequential Sampling

تستخدم هذه الطريقة بشكل واسع في رقابة الجودة ويمكن أن تكون مفيدة في دراسات الميزانية الزراعية. حيث تشتمل على اختبار للعينات الصغيرة نسبياً في رقابة الجودة، بناء على قاعدة مخرجات العينة، واتخاذ القرار بقبول أو رفض كامل المنتج. وإذا لم تؤدِ العينة الصغيرة إلى قرار واضح يمكن زيادة حجم العينة. فمثلاً يتم سحب وحدات إضافية حتى نستطيع الوصول للقرار. هذه الطريقة تحافظ على انخفاض تكاليف المعاينة؛ نظراً لأنه في معظم الأحيان يمكننا صنع القرار بناء على النتائج المتحصل عليها من العينات الصغيرة جداً. وفي دراسات الميزانية، يتم مقابلة مجموعات صغيرة من المزارعين بخصوص تطبيقات الإنتاج. فإذا كانوا موافقين يتم تحديد الميزانية بناء على تلك التطبيقات أما إذا كان خلاف ذلك يتم أخذ عينه أكبر حجم حتى نصل إلى اتفاق.

العينات غير الاحتمالية Non-probability Samples

تتميز طريقه العينات غير الاحتمالية بأن جميع مفردات المجتمع ليس لها فرصة متساوية في الاختيار. ويتم استخدام هذه الطريقة بسبب الكفاءة والتكلفة ولكنها لا تعطي نتائج موضوعية؛ كما في طريقة العينات الاحتمالية.

المعاينة الملائمة Convenience Sampling

باستخدام طريقه المعاينة الملائمة يتم اختيار العينة بناء على ما هو مناسب لنا بدون وضع اعتبار لتمثيل المجتمع بصورة ملائمة. وتتمثل ملائمتها في الوقت والتكاليف ورأي الإدارة. ونرى ذلك النوع من ا لمعاينة في التلفزيون المحلي عندما يكون التقرير المقدّم عبارة عن مقابلة مع شخص في الشارع للحصول على رأيه حيال بعض المواضيع في الأخبار. وهذا التصميم يجب أن لا يستخدم لعمل استدلال عن المجتمع.

المعاينة التحكمية (الاجتهادية) Judgment Sampling

في هذه الحالة يتم استخدام الاجتهاد الشخصي بدلاً من بعض الطرق الموضوعية مثل الأرقام العشوائية لاختيار العناصر من المجتمع الممثلة للعينة. فمثلاً ، وكيل التعليم ا لمحلي ربما يتم اختياره بواسطة الباحث لاختيار المزارعين يدوياً لمقابلتهم لمعرفه رأيهم حيال التطبيقات الإدارية ؛ نظراً لأن الباحث يرغب في التحدث فقط لأفضل المديرين.

المعاينة الحصصية Quota Sampling

في المعاينة الحصصية يتم تصميم العينة لتكون مشابهة للمجتمع تبعاً لبعض الصفات الرئيسية. ويتم سحب العناصر من المجتمع حتى نصل لحصة تلك الصفة ثم يتم اختيار عناصر تحتوي على صفة مختلفة وهكذا. فمثلاً إذا كان لدينا معرفة بأن المحافظة أو المقاطعة تحتوي ٢٠ ٪ مزارعين قمح ، ٦٠ ٪ مزارعين قطن و٢٠ ٪ مربين للثروة الحيوانية فإنه يمكن اختيار عينة حصصية تشتمل على نفس النسب. ونظراً لأن تقديرات هذا النوع من العينات معرض لاختلافات كبيره مقارنة بالمعاينة الاحتمالية فإنه يفضل استخدام طرق المعاينة الاحتمالية.

توزيعات المعاينة Sampling Distributions

يساعدنا توزيع المعاينة لفهم المشاكل المتعلقة بالمعاينة. حيث يمكننا باستخدام توزيع المعاينة معرفة اختلاف الإحصاءات مثل المتوسط والتباين من عينه إلى عينه عند سحب عينات عشوائية متتالية بنفس الحجم ومن نفس المجتمع ويسمى التوزيع الاحتمالي لمثل هذه الإحصاءات بتوزيع المعاينة. ولذلك فإنه سيكون هناك توزيع معاينه للمتوسط وللتباين وهكذا. وتعبّر هذه التوزيعات عن المفهوم الأساسي للاستدلال الإحصائي وهي كذلك مهمة في نظرية القرار الإحصائي.

توزيع المعاينة للمتوسط الحسابي Sampling Distribution of the Mean

تعتمد فكرة توزيع المعاينة على المعاينة المتكررة. ولذلك عند سحب مجموعة من العينات بنفس الحجم من مجتمع معين وحساب المتوسط الحسابي لكل عينه ، فإنه يمكن معاملة تلك المتوسطات كمتغير عشوائي. ويعني ذلك أن المتوسط الحسابي يأخذ قيماً متغيرة بناء على العناصر التي تم سحبها من المجتمع بطريقة عشوائية كعينة. وعند معاملة جميع العينات العشوائية باحتمالات متساوية ، فإنه يمكن وضع قيم مختلفة للمتوسط الحسابي في شكل توزيع احتمالي عن طريق جمع الاحتمالات المرتبطة بكل قيمة. التوزيع الاحتمالي الناتج يتميز بشكل معين بغض النظر عن شكل التوزيع الاحتمالي للمجتمع الذي تم سحب العينات منه. بالإضافة لذلك فإن القيمة المتوقعة للتوزيع الاحتمالي تساوي قيمة متوسط المجتمع وكذلك تباين التوزيع الاحتمالي ذو علاقة بتباين المجتمع. ويمكن إيضاح ذلك من خلال المثال التالي

نفرض أن لدى مزارع مجموعة عجول صغيرة عمرها سنة واحدة وعددها ستة وكانت أوزانها عند الولادة كالتالي (رطل)

$$X = \{90, 80, 100, 80, 90, 100\}$$

وعليه فإن متوسط المجتمع آنفاً يساوي ٩٠ رطلاً وتباينه σ^2 يساوي ٦٦٫٦٧. وفي حالة الرغبة في اختيار عينة من ذلك المجتمع حجمها ٢ عجل بدون إرجاع فإن عدد العينات الممكن الحصول عليها يمكن حسابه كالتالي :

$$_6C_2 = 6!/(6-2)!(2!) = 15$$

و يمكن حساب المتوسط لكل عينه واستخدام ذلك المتوسط كمتغير عشوائي لإيجاد توزيع المعاينة للمتوسط (الجدول رقم (٥٫١).

ويتضح من الجدول أن القيم المحتملة للمتوسط قريبة جداً من متوسط المجتمع. ونظراً لوجود تكرار في قيم المتوسط للعينات المختارة فإنه يمكن إيجاد التوزيع الاحتمالي لقيم المتوسط كما في الجدول رقم (٥٫٢). حيث يمكن حساب متوسط التوزيع الاحتمالي في الجدول باستخدام صيغة القيمة المتوقعة كما يلي :

$$E(\overline{X}) = \Sigma \overline{X} \cdot P(\overline{X}) = (80)(1/15) + (85)(4/15) + (90)(15/15) + (95)(4/15) + (100)(1/15) = 1350/15 = 90$$

والتي تساوي نفس القيمة التي تم حسابها لمتوسط المجتمع. وهذه العلاقة متحققة دائماً طبقاً لمفهوم نظرية النهاية المركزية.

نظرية النهاية المركزية Central Limit Theorem

إذا تم اختيار عينات عشوائية بحجم n مشاهدات من مجتمع له متوسط منتهي μ وانحراف معياري σ ، فإن توزيع متوسط العينة $\overline{X}$ يقترب من التوزيع الطبيعي بمتوسط μ وانحراف معياري $\frac{\sigma}{\sqrt{n}}$ عندما يكون حجم العينة n كبير. وكلما زاد حجم العينة n كان التقريب أكثر دقة.

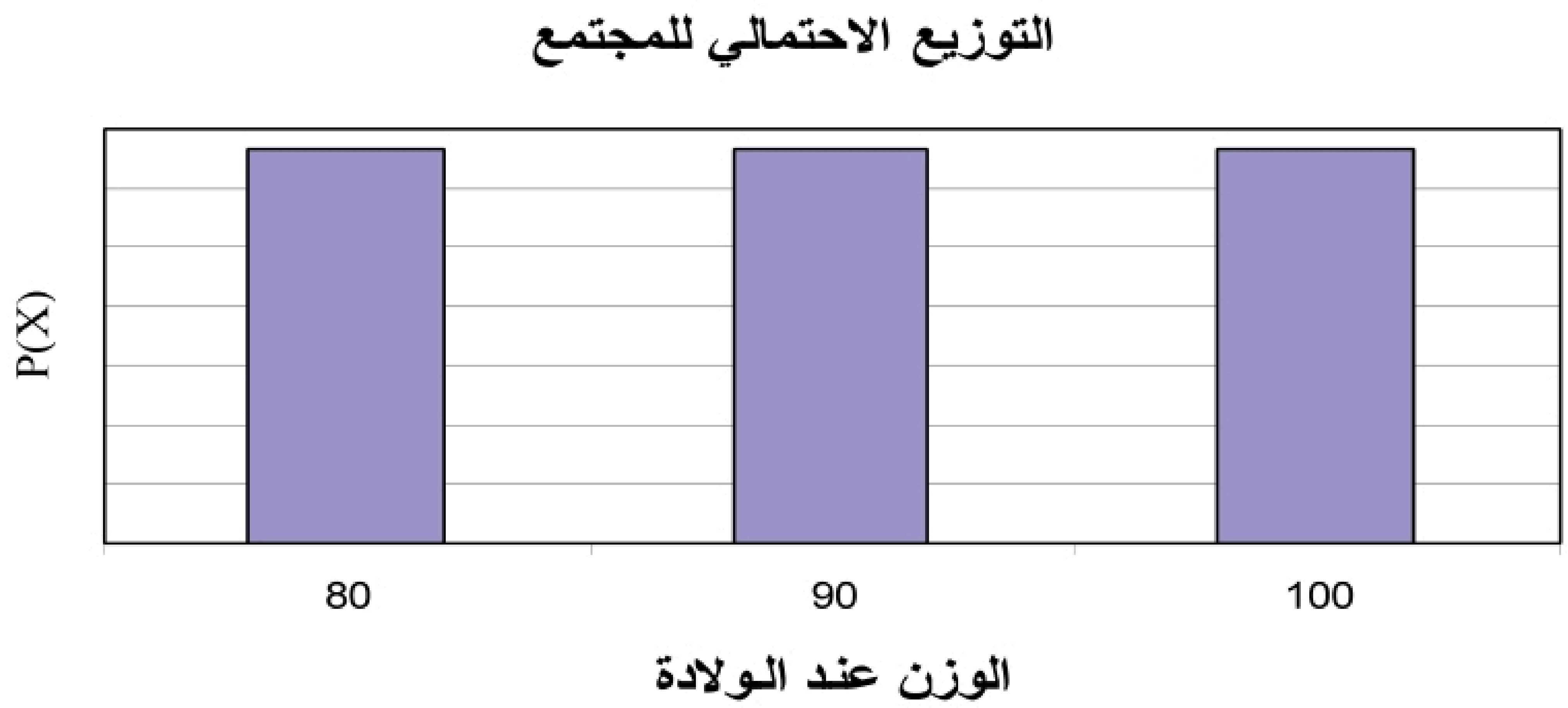
وعند رسم التوزيع الاحتمالي لمجتمع صغير من أوزان مواليد العجول فإنه يأخذ شكل المستطيل كما في الشكل رقم (٥,١)، ولكن توزيع العينة للمتوسط والذي تم إيجاده برسم جميع قيم المتوسطات الممكنة واحتمالاتها المصاحبة يأخذ شكل التوزيع الطبيعي، كما في الشكل رقم (٥,٢).

الجدول رقم (٥,١). متوسطات العينة لجميع التوليفات الممكنة لعينة حجمها اثنين.

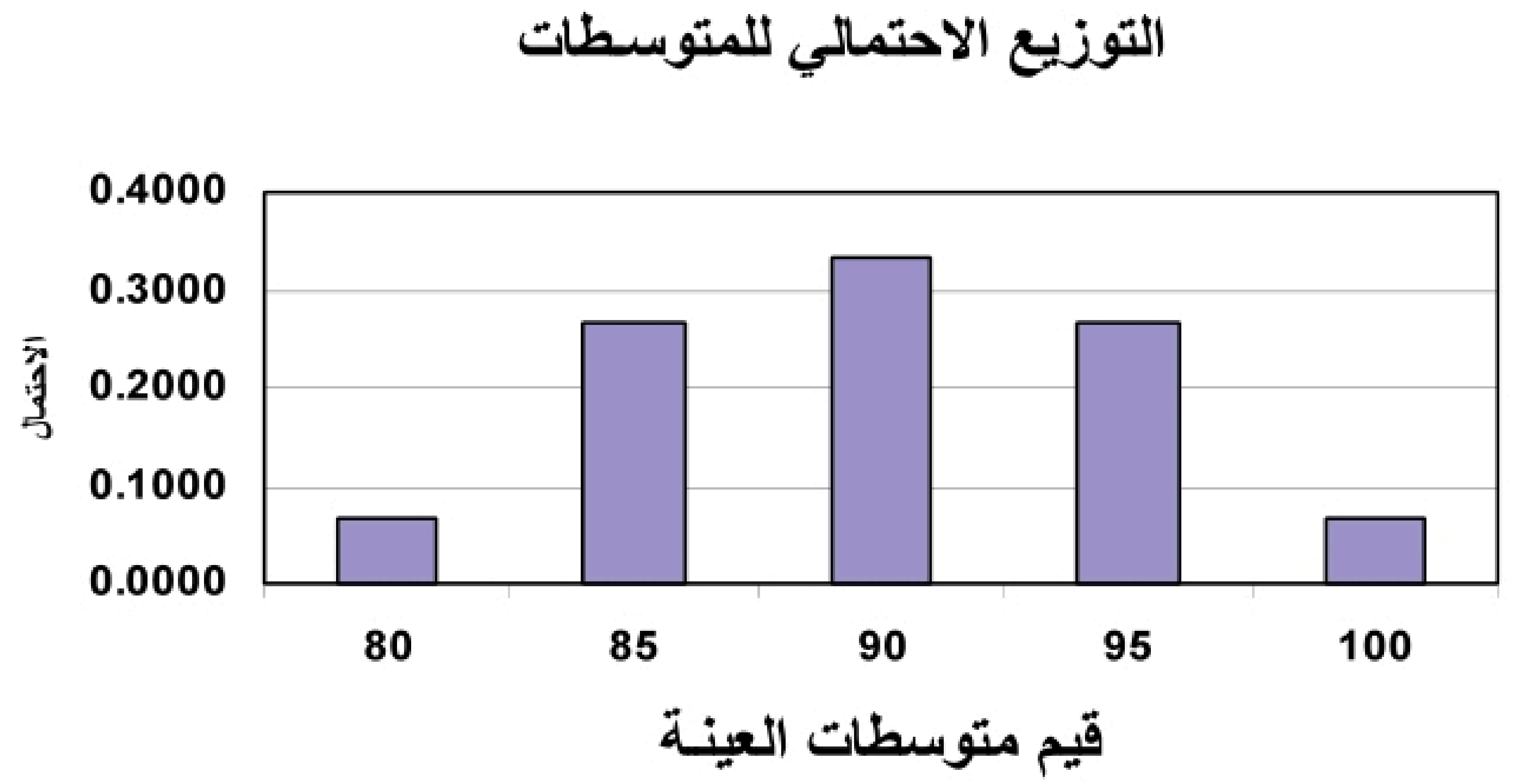
توليفة العينة	قياسات العينة	المجموع (ΣX)	متوسط العينة ($\bar{X}$)
X_1,X_2	٩٠، ٨٠	١٧٠	٨٥
X_1,X_3	٩٠، ١٠٠	١٩٠	٩٥
X_1,X_4	٩٠، ٨٠	١٧٠	٨٥
X_1,X_5	٩٠، ٩٠	١٨٠	٩٠
X_1,X_6	٩٠، ١٠٠	١٩٠	٩٥
X_2,X_3	٨٠، ١٠٠	١٨٠	٩٠
X_2,X_4	٨٠، ٨٠	١٦٠	٨٠
X_2,X_5	٨٠، ٩٠	١٧٠	٨٥
X_2,X_6	٨٠، ١٠٠	١٨٠	٩٠
X_3,X_4	١٠٠، ٨٠	١٨٠	٩٠
X_3,X_5	١٠٠، ٩٠	١٩٠	٩٥
X_3,X_6	١٠٠، ١٠٠	٢٠٠	١٠٠
X_4,X_5	٨٠، ٩٠	١٧٠	٨٥
X_4,X_6	٨٠، ١٠٠	١٨٠	٩٠
X_5,X_6	٩٠، ١٠٠	١٩٠	٩٥

الجدول رقم (٥,٢). التوزيع الاحتمالي لمتوسطات العينة.

متوسط العينة ($\bar{X}$)	التكرار (f)	الاحتمال، P ($\bar{X}$)	$\bar{X}$ .P($\bar{X}$)
٨٠	١	١٥/١	١٥/٨٠
٨٥	٤	١٥/٤	١٥/٣٤٠
٩٠	٥	١٥/٥	١٥/٤٥٠
٩٥	٤	١٥/٤	١٥/٣٨٠
١٠٠	١	١٥/١	١٥/١٠٠
الإجمالي	١٥	١=١٥/١٥	٩٠=١٥/١٣٥٠



الشكل رقم (٥,١). التوزيع الاحتمالي لمجتمع أوزان العجول المكتسبة.



الشكل رقم (٥,٢). توزيع المعاينة لمتوسطات العينة لعينات حجمها يساوي اثنين.

الخطأ المعياري للمتوسط Standard Error of the Mean

الانحراف المعياري لتوزيع المعاينة للمتوسط يسمى الخطأ المعياري للمتوسط وبذلك فإنه ليس هناك خلط بينه وبين الانحراف المعياري للمجتمع. الخطأ المعياري مرتبط بالانحراف المعياري للمجتمع ؛ كما في نظرية النهاية المركزية فهو يساوي حاصل قسمة الانحراف المعياري للمجتمع على الجذر التربيعي لحجم العينة.

ولكن عندما يكون لدينا مجتمعاً صغيراً كما في مثال أوزان مواليد العجول ، فإنه يجب استخدام معامل تصحيح المجتمع المحدود. وتوضح المعادلة التالية رقم (5.1) طريقة حساب الخطأ المعياري للمتوسط باستخدام معامل التصحيح :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad (5.1)$$

حيث :

σ = الانحراف المعياري للمجتمع.

N = حجم المجتمع.

n = حجم العينة.

$\sqrt{\frac{N-n}{n-1}}$ = معامل تصحيح المجتمع المحدود.

وبتطبيق الصيغة الرياضية (5.1) على مثال أوزان مواليد العجول نحصل على :

$$\sigma_{\bar{x}} = \frac{8\cdot 167}{\sqrt{2}}\sqrt{\frac{6-2}{6-1}} = (5\cdot 78)\ (0.8944) = 5.17$$

حيث إن معامل التصحيح يساوي تقريباً ٠,٨٩ ولذلك عمل على تخفيض $\sigma_{\bar{x}}$ نوعاً ما.

الجدول رقم (٥.٣). حساب الخطأ المعياري للمتوسط باستخدام التوزيعات الاحتمالية.

متوسط العينة، $\bar{X}$	$P(\bar{X})$	$\bar{X}$ - μ	$(\bar{X}$ - μ$)^2$	$(\bar{X}$ - μ$)^2 . P(\bar{X})$
٨٠	١/١٥	٨٠-٩٠=-١٠	١٠٠	١٠٠/١٥
٨٥	٤/١٥	٨٥-٩٠=-٥	٢٥	١٠٠/١٥
٩٠	٥/١٥	٩٠-٩٠=٠	٠	٠
٩٥	٤/١٥	٩٥-٩٠=٥	٢٥	١٠٠/١٥
١٠٠	١/١٥	١٠٠-٩٠=١٠	١٠٠	١٠٠/١٥
الإجمالي	١٥/١٥=١			$\sigma^2_{\bar{x}}$ =400/15=26.67
				$\sigma_{\bar{x}} = \sqrt{26.67} = 5.17$

بطريقه أخرى ، يمكن حساب $\sigma_{\bar{X}}$ مباشره من البيانات المعطاة في الجدول رقم (٥,٢). وذلك باستخدام الصيغة الرياضية لحساب التباين للتوزيع الاحتمالي للبيانات التالية :

$$\sigma^2 = \Sigma(X-\mu)^2 \cdot P(X) \qquad (5.2)$$

وتشير النتائج المتحصل عليها في الجدول رقم (٥,٣) إلى الوصول لنفس النتيجة التي حسبت سابقاً $\sigma_{\bar{x}} = 5.17$. ولكن في حالة معرفتنا لقيمة الانحراف المعياري للمجتمع فإننا نستخدم طريقه معامل التصحيح لسهولتها. وفي الحقيقة ، فإنه لا يتم سحب جميع العينات الممكنة من أي توزيع حيث يتم في العادة سحب عينة واحدة. ولذلك فإنه ليس لدينا الخيار لحساب $\sigma_{\bar{x}}$ باستخدام متوسطات التوزيعات الاحتمالية كما في الجدول رقم (٥,٣) ، ولحسن الحظ فإن لدينا صيغة رياضيه معتمده على الانحراف المعياري للمجتمع σ .

تمارين Exercises

١- ما الفرق بين التعداد والعينة؟ اشرح مع إعطاء ثلاثة أسباب أو أكثر لماذا المعاينة أكثر ملائمة من التعداد في حالات معينة.

٢- أعط مثال للحالات التي يكون فيها المجتمع :

أ) غير محدود.

ب) يعتبر غير محدود ، ولكنه في الحقيقة محدود.

ج) محدود.

٣- ما الهدف من وجود مجموعه ضابطة في الدراسة؟ أعط مثال لحالة تكون فيها المجموعة الضابطة ذات أهمية.

٤- ما الفرق بين الأخطاء العشوائية والمنتظمة؟ وما الذي يجب تقليله في المعاينة؟ اشرح.

٥- باستخدام جدول الأرقام العشوائية، الملحق جدول (١)، اختر عينة عشوائية بسيطة حجمها عشر مفردات من مجتمع حجمه ١٠٠ مع شرح خطوات الاختيار المستخدمة.

٦- باستخدام جدول الأرقام العشوائية، الملحق جدول (١)، اختر عينه عشوائية منتظمة حجمها ٥ ٪ من مجتمع حجمه ٣٠٠. وضح خطوات الاختيار المستخدمة.

٧- أعد حل تمرين ٥ ، ٦ باستخدام أسلوب المعاينة من قائمة أدوات : تحليل البيانات في برنامج الجداول الالكترونية Excel . أطبع النتائج التي تم الحصول عليها ثم اشرح خطوات الاختيار المستخدمة.

٨- لدى شركة الزراعة العالمية مندوبي مبيعات في سكرامنتو ومينابليس، هيوستن، وأتلانتا. ويحتوي كل مكتب على ١٦ مندوباً. فإذا كانت المبيعات الأسبوعية لأي مندوب مبيعات تتوزع طبيعياً بمتوسط ٣٠,٠٠٠ دولار وانحراف معياري ٦٤٠٠ دولار فما هو المدى حول المتوسط بنسبه ٩٥,٥٪ للحالات التالية:

أ) المبيعات الأسبوعية لأي مندوب مبيعات تم اختياره عشوائياً.

ب) متوسط المبيعات الأسبوعية للمندوب في مكتب هيوستن.

ج) متوسط المبيعات الأسبوعية للمندوب في الشركة.

٩- ما هو احتمال اختيار عينة عشوائية بسيطة بمتوسط ٥٠ أو أكثر من مجتمع متوسطه ٤٦ إذا كان حجم العينة ٨١ وتباين المجتمع ٣٢٤؟

١٠- إذا كان متوسط المشتريات الأسبوعية من الطعام لعدد ٥٠ مستهلك هي ١٢٠ دولاراً بانحراف معياري ٢٥ دولار. تم اختيار مجموعة عشوائية مكونة من ٥ أشخاص من هذا المجتمع لدراسة عاداتهم الشرائية. المطلوب إيجاد متوسط مشتريات هذه المجموعة من الطعام والانحراف المعياري.

١١- معمل للأصول الوراثية لديه ٥ مواقع مزروعة بأصناف حديثه من القطن لها إنتاجية مختلفة كالتالي :

الموقع	إنتاجية الايكر
V	٥٢٠
W	٥٠٠
X	٥٤٠
Y	٥٧٠
Z	٤٨٠

تم اختيار موقعين عشوائياً لتقدير متوسط الإنتاجية ، المطلوب :

أ) أوجد جميع العينات الممكنة المكونة من موقعين ثم احسب متوسط الإنتاجية لكل عينه.

ب) احسب المتوسط والتباين للتوزيع الاحتمالي من البيانات في الفقرة أ وكذلك من البيانات الأصلية.

١٠- قررت مجموعة كبيرة من مزارعي البرتقال في الشاطئ الغربي - Cal co-op إنشاء مصنع للعصير. تم اختيار عدد ١٦ مصنعاً متشابهه بطريقه عشوائية وقدّر متوسط إنتاجها فكان٣٠٠٠ جالون عصير في اليوم بانحراف معياري ٤٠٠ جالون. ما هو احتمال أن يكون للمصنع المختار عشوائياً متوسط يومي في حدود ٢٠٠ جالون من متوسط المجموعة.

الفصل السادس

مقدمة في الاستدلال الإحصائي: عينة واحدة

Introduction to Statistical Inference: one Sample

في علم الإحصاء نتعامل غالباً مع العينات، على الرغم أن اهتمامنا المبدئي متعلق بالمجتمع الذي تم اختيار هذه العينات منه. ويتمثل الهدف في إيجاد أو تقدير بعض قيم متغيرات المجتمع اعتماداً على نتائج العينة. وكما تمت الإشارة إليه سابقاً، فإن العينة العشوائية هي في الحقيقة مجموعة جزئية عشوائية من المجتمع ويتركز اهتمامنا في عمل استدلال حول المجتمع. ولكن المشكلة تكمن في عدم التأكد من صحة الاستدلال؛ نظراً للتباين في البيانات المعتمد عليها علاوة على التباين بين العينات المختارة. ولتقليل تلك الأخطاء، فإننا نستخدم نظرية الاحتمال كقاعدة للاستدلال. ولذلك فإنه يمكن تعريف الاستدلال الاحصائي بأنه تقدير لبعض المتغيرات في المجتمع بناء على نظرية الاحتمال وبيانات العينة.

يشتمل الاستدلال الإحصائي على تقدير متغيرات المجتمع واختبارات الفروض حول تلك الظواهر. وتسمى متغيرات المجتمع معالم parameters ومتغيرات العينة تسمى إحصاءات Statistics. وبذلك، فإنه غالباً يتم حساب الإحصاءات من بيانات العينة ثم نستخدمها كقاعدة لتقدير معالم المجتمع. وبصفة عامة، فإن المتغيرات العشوائيه التي تم

اختيارها لتقدير معالم المجتمع تسمى مقدرات estimators ، في حين أن القيم المحددة لهذه المتغيرات العشوائية تسمى تقدير estimates لمعالم المجتمع. ويمكن عرض هذه التقديرات بطريقتين مختلفتين هما التقدير بنقطه والتقدير بفترة. حيث إن التقدير بنقطة عبارة عن قيمة واحدة حيث يمكن تقدير متوسط المجتمع من بيانات العينة ويساوي مثلاً ٢. في حين أن التقدير بفترة يساوي مثلاً ١-٣ بحيث إن متوسط المجتمع يقع داخل هذه الفتره. ويعتمد اختيار أحد هذه الطريقتين على الحالة موضع الدراسة علاوة على تقدير الاحتمال المصاحب لها.

خصائص المقدرات Properties of Estimators

متوسط وتباين المجتمع هي المعالم التي يتم تقديرها غالباً ؛ لأنها تحدد التوزيعات الاحتماليه. فعند الحاجة لتقدير متوسط المجتمع فإن السؤال يتمثل في اختيار الاحصاءت المناسبة للتقدير. على الرغم من أن هناك العديد من المقدرات الممكن الاختيار منها الاّ أن أفضلها هو متوسط العينة $\overline{X}$. وذلك نظراً لتحقيقها لمجموعة من الخصائص التي يجب أن يتميز بها المقدر الجيّد والتي يمكن عرضها كالتالي :

عدم التحيز Unbiasedness

يكون المقدر غير متحيز إذا كانت القيمة المتوقعه لمتوسط العينة مساوية لمعلمة المجتمع ، $E(\overline{X}) = \mu$. أو بصفه عامه ، إذا كانت $\hat{\eta}$ [الحرف الإغريقي η مضاف له ^ (هات)] وتقرأ (إيتا هات) تمثل المقدر أو إحصاء العينة و η هي معلمة المجتمع فإن $\hat{\eta}$ مقدر غير متحيز إذا كانت $E(\hat{\eta}) = \eta$ ، مثلاً إذا كان توزيع المعاينة لها يتمركز مباشرةً على η. ولذلك فإن $\hat{\eta}$ تكون متحيزة إذا كانت قيمتها المتوقعه تختلف عن η.

ويمكن تعريف التحيز B بمعادلة رياضيه كالتالي :

$$B \equiv E(\hat{\eta}) - \eta \tag{6.1}$$

هذا يعني أن المقدر الجيد يجب أن يكون غير متحيز، ويشار إلى أن $\overline{X}$ (متوسط العينة) غير متحيز ؛ لأننا سبق وأن عرفنا أن متوسط توزيع المعاينة $\overline{X}$ مساوٍ لمتوسط المجتمع μ عند مناقشة توزيع المعاينة للمتوسط.

الكفاءة Efficiency

كما أننا نهدف لأن يكون المتوسط ذو كفاءة فنحن أيضاً نرغب بأن يكون لتوزيع المعاينة للمقدر أقل تباين وهذا يحقق الكفاءة. وبذلك إذا كان لدينا مقدرين $\hat{\eta}$ ، η' (إيتا برايم) للمتوسط η فإن المقدر $\hat{\eta}$ أكبر كفاءة من المقدر η' إذا كان التباين لتوزيع المعاينة لـ $\hat{\eta}$ أقل. أي أن توزيع المعاينه لها أكثر تدبّباً. وعليه فإن المقدره $\hat{\eta}$ الأكثر كفاءة سوف تعطي تقدير نقطة قريب من هدفنا المتمثل في القيمة الصحيحة للمقدر η. أو أننا سنعطي فتره أصغر في حالة تقدير الفتره ولذلك تكون أكثر دقة لنفس الحجم من العينة.

وبالطبع فإن زيادة حجم العينة n سيؤدي لتقليل التباين لبعض المقدرات. وعليه فإن الطريقه الأخرى للحصول على كفاءة أعلى لـ $\hat{\eta}$ هي معرفة أن العينات الكبيره سوف تحقق ذلك سواء على مستوى تقدير النقطه أو الفتره وعليه فإن استخدام $\hat{\eta}$ يحقق كفاءة أكثر نظراً لأن تكاليف العينة صغيره. من جهة أخرى فإنه يمكن إيجاد مقياس نسبي مناسب للكفاءة لمقدرين غير متحيزة بحساب نسبة التباين لهما، كما في المعادلة التالية :

$$\text{الكفاءة النسبية لـ } \hat{\eta} \text{ مقارنة بـ } \eta' = var(\eta') / var(\hat{\eta}) \tag{6.2}$$

الكفاية Sufficiency

خاصيتي عدم التميز والكفاءة مهمة للمقدرات خاصة في حالة العينات الصغيرة.

الخاصيه الأخرى المرغوبة هي الكفايه حيث إن المقدر الذي يحقق الكفاية $\hat{\eta}$ يستخدم جميع المعلومات المتاحة حول معلمة المجتمع η المشمولة في بيانات العينة. وعليه فإن المقدر الكافي يستخدم جميع مشاهدات العينة في الحسابات. فمثلاً الوسيط Md لايحقق شرط الكفاية حيث إنه يستخدم ترتيب المشاهدات وليس قيمها ، ولكن المتوسط $\overline{X}$ والتباين S^2 عبارة عن مقدرات تحقق الكفاية. ولذا فإن الكفاية شرط ضروري لتحقيق الكفاءة.

الاتساق Consistency

المقدر المتسق $\hat{\eta}$ هو الذي ينطبق تماماً على مقدر المجتمع كلما زاد حجم العينه ليصل إلى مالا نهاية. ولذلك عندما يقترب حجم العينة لما لانهاية فإن المقدر المتسق $\hat{\eta}$ يعطي تقدير نقطه مطابق تماماً لمقدر المجتمع η. وكما أن التباين يعتبر مقياساً جيّداً لانتشار التوزيع حول متوسطه فإن متوسط مربع الخطأ (MSE) مقياس جيد لانتشار المقدر $\hat{\eta}$ حول القيمة الحقيقيه η في توزيع المعاينة. ولذلك فإن الاتساق يعني أن متوسط مربعات الخطأ يساوي صفر عندما يصل حجم العينة لما لانهاية والذي يمكن إيضاحه من خلال المعادلة التالية :

$$MSE \equiv E(\hat{\eta} - \eta)^2 \tag{6.3}$$

و $\hat{\eta}$ متسق كلما اقتربت حجم العينه من ما لانهايه $\boldsymbol{n} \to \infty$ و $MSE \to 0$

وتوضح المعادلة رقم (6.4) العلاقة بين متوسط مربع الخطأ (MSE) والتحيز والتباين. ولذلك فإن $\hat{\eta}$ مقدر متسق إذا فقط كان تحيزه وتباينه تؤول للصفر كلما اقترب حجم العينة من ما لانهاية.

$$MSE \equiv E(\hat{\eta} - \eta)^2 = \left[Bias\ \hat{\eta}\right]^2 + var(\hat{\eta}) \qquad (6.4)$$

وفي حالة أن التحيز، اقترب من الصفر فإن المقدر يدعى مقدر غير متحيز تقريباً بحيث إنه حقق الشرط الأضعف للاتساق.

وبصفه عامة فإن خاصية الاتساق للمقدر لا تضمن أن المقدر هو الأفضل. فمثلاً، في حالة المقدر μ في مجتمع طبيعي، يكون وسيط العينه Md متسقاً؛ نظراً لأنه غير متحيز وتباينه يقترب من الصفر كلما زاد حجم العينة يقترب من ما لانهاية ولكن ليس بنفس كفاءة متوسط العينة $\overline{X}$ لأن متوسط العينه $\overline{X}$ يحقق الاتساق والكفاءة معاً، فمثلاً تباين المتوسط دائماً أقل من تباين الوسيط. ولذلك فإننا نهدف لايجاد مقدر يحقق جميع تلك الخصائص، والتي تتحقق في حالة متوسط العينه $\overline{X}$ وتباينها S^2. وعليه فإن هذين المقدرين هما المقدرات المعتاد استخدامها كتقدير نقطه لمتوسط المجتمع وتباينه μ ، σ^2.

تقدير الفترة Interval Estimation

ليس بالضرورة أن يكون تقدير معالم المجتمع قيمة واحدة، حيث يمكن أن يكون عبارة عن مدى من القيم. وتسمى التقديرات التي تشتمل على مدى من القيم بتقديرات الفتره أو فترات الثقة.

ويتم بناء فترات الثقة لمعالم المجتمع مثل المتوسط μ بناءً على توزيع معاينة صحيح للمقدر مثل توزيع المعاينه لمتوسط العينه $\overline{X}$. فمثلاً في حالة استخدام متوسط العينه $\overline{X}$ لعمل استدلال لمتوسط المجتمع μ فإنه يجب إضافة مقياس احتمالي للفتره وهذا يعني أنه يمكن القول بأنه عند الرغبة في بناء فتره ثقة لقيم المتوسط $\overline{X}$ المتحصل عليها من عينات مكررة حجمها n مسحوبة من المجتمع فإننا متأكدين بنسبة

$100(1-\alpha)$ أنها تحتوي على متوسط المجتمع μ.

وقد تم استخدام مقياس احتمالي ($1-\alpha$) حيث α (تشير إلى حرف إغريقي) عبارة عن قيم صغيره جداً وفي العادة تكون قيمته أقل من ٠.١ ويتم اختياره بناء على مقدار تكاليف الاختيار الخاطئ. فكلما كانت التكلفه أكبر كان الرقم أصغر. ولذلك فعند اختيار $\alpha = 0.05$ فإن فتره الثقة هي ٩٥٪.

ويجب ملاحظة أنه في تفسيرنا لفتره الثقة، تمت الاشارة إلى عينات مكررة بحجم n. ولكن في التطبيق العملي يتم اختيار عينة واحدة ونحسب قيمة المقدر (مثل $\overline{X}$) ثم نختار درجة الثقة (اختيار $1-\alpha$) ، ثم يتم بناء فتره الثقة ويتم ذلك على اعتبار أن متوسط المجتمع μ مشمول في العينة المختارة. وفي الحقيقه فإن ذلك منطقي ويكون المقدر ضمن الفتره ما لم تكن القيمة المختارة واقعة في أحد النهايات الطرفية لتوزيع المعاينة. وكما هو معلوم فإن احتمال الحصول على قيمة تقع في أحد الأطراف لتوزيع المعاينة يساوي α والتي تم اختيارها لتكون قمية صغيرة.

من جهة أخرى فإن طول فترة الثقة يعتمد على القيم المختارة لحجم العينة n و مستوى المعنوية α . وكلما كان بحجم العينة صغير كانت الفترة أطول نظراً لأن تغيرات توزيع المعاينة مرتبط بحجم العينة n $(\sigma_{\bar{x}} = \sigma/\sqrt{n})$.وكلما كانت الاختلافات في توزيع المعاينة كبيره كلما كان التوزيع أوسع انتشاراً، ومن ثمَّ فتره ثقة أطول. أيضاً فإن اختيار قيمه صغيره لمستوى المعنوية α يعطي فترة ثقة أطول. وكلما كانت الفترة أطول، كان هناك مجال أكبر للمقدرات. وعليه فإنه يجب الموازنة بين تكاليف مشاهدات العينة الإضافيه وتكاليف الحصول على فترة ثقة واسعة.

والعامل الأخير المؤثر في حجم فترة ثقة معيّنة هو توزيع المعاينة للإحصاء المستخدم لتقدير المعلمه. وعموماً فإن طريقة الحصول على فتره الثقة للمعلمة η هو

إيجاد تقدير نقطه أولاً للمعلمة وهو $\hat{\eta}$. ثم بعد ذلك بناء فترة الثقة بإيجاد قيم الفترة حول $\hat{\eta}$ المبنية على توزيع المعاينة ، وبالتالي الحصول على درجه الثقة المرغوبه. ونظراً لاستخدامنا لتوزيعات معاينة مختلفة لتقدير متوسط المجتمع ، مقدرة ذي الحدين p ، وتباين المجتمع فإننا سنقوم بوصف طريقه بناء فترات الثقة لكل منهما في الموضوع التالي.

فترات الثقة لمتوسط مجتمع μ انحرافه المعياري σ معلوم

Confidence Intervals for μ with σ Known

لبناء فترة الثقة لمتوسط المجتمع ، نفترض أننا سحبنا عينة عشوائية من مجتمع يتوزع طبيعياً له انحراف معياري σ معلوم. وعليه فإن إحصاء العينة للمتوسط المقدر μ هو $\overline{X}$ ؛ نظراً لشمولها لجميع الخصائص المرغوبة للمقدر التي تمت مناقشتها سابقاً. وتوزيع المعاينة للمتوسط $\overline{X}$ يتوزع طبيعياً بمتوسط $E(\overline{X}) = \mu$ وخطأ معياري $\sigma_{\overline{x}} = \sigma / \sqrt{n}$.

والمتغير z والموضح في المعادله رقم (6.5) عبارة عن توزيع طبيعي قياسي له متوسط يساوي الصفر وانحراف معياري يساوي الواحد $Z \sim N(0,1)$ ويمكن إيجاد الاحتمالات للتوزيع Z باستخدام الجدول رقم (٧) المرفق في الملاحق.

$$z = (\overline{X} - \mu) / \sigma_{\overline{X}} \qquad (6.5)$$

الآن تفترض أننا عرفنا Z_α بحيث تعبّر عن قيمة z الواقعة في الطرف الأيمن للتوزيع بحيث أن احتمال ملاحظة قيم لـ z أكبر من z_α يساوي α ، أي $P(Z > Z_\alpha) = \alpha$. وأيضاً $-z_\alpha$ عبارة عن مكان في الطرف الأيسر في التوزيع بحيث $P(Z < -Z_\alpha) = \alpha$ كما هو موضح في الشكل رقم (٦،١). وبدلاً من ايجاد موقعين

لتحديد مكان α في طرف التوزيع الطبيعي القياسي كان من المناسب ايجادها لـ $\alpha/2$ في كل طرف (بحيث يكون مجموع المساحة مساوٍ لـ α). ويمكن إيضاح ذلك بحيث أن $Z_{\alpha/2}$ تمثل القيمة الاحتمالية $P(Z > Z_{\alpha/2}) = \alpha/2$ و $Z_{\alpha/2}$ - تمثل القيمة الاحتماليه $P(Z < -Z_{\alpha/2}) = \alpha/2$. فمثلاً إذا كانت $\alpha = 0.05$ فإن $\alpha/2 = 0.025$ وبذلك قيمة $Z_{\alpha/2}$ والتي تكون المساحة الواقعه على يمينها مساوية 0.025 هي ١,٩٦ (من الجدول ٧ في الملحق) ونظراً لأن التوزيع الطبيعي متماثل فإن قميه $Z_{\alpha/2}$ - يجب أن تساوي - ١,٩٦ ولذلك تكون قيمة الاحتمال أن Z تقع بين القيمة ±١,٩٦ تساوي :

$$P(-1.96 < Z < 1.96) = (1.0 - 0.05) = 0.95$$

وهي التي نرغبها لفترة ثقة ٩٥٪. ولذا، بصفه عامه فإن الاحتمال $(1-\alpha)$ معطى بالمعادلة (6.6) وهو يمثل $100\,(1-\alpha)$ لفترة ثقة لـ Z .

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = (1-\alpha) \qquad (6.6)$$

ولكن المطلوب هو إيجاد فترة الثقة لمتوسط المجتمع μ وليس لـ Z . ويمكننا عمل ذلك عن طريق معالجة المعادلة رقم (6.6) رياضياً. حيث نعوّض عن قيمة Z في المعادلة بقيمة Z الموضحة في المعادلة رقم (6.5) ونعيد ترتيب المعادلة للحصول على فترة الثقة للمعلمة μ ويمكن إيضاح ذلك كالتالي :

$$P(-Z_{\alpha/2} < (\overline{X} - \mu) / \sigma_{\overline{X}} < Z_{\alpha/2}) = (1-\alpha)$$

وبالضرب في $\sigma_{\bar{x}}$ نحصل على :

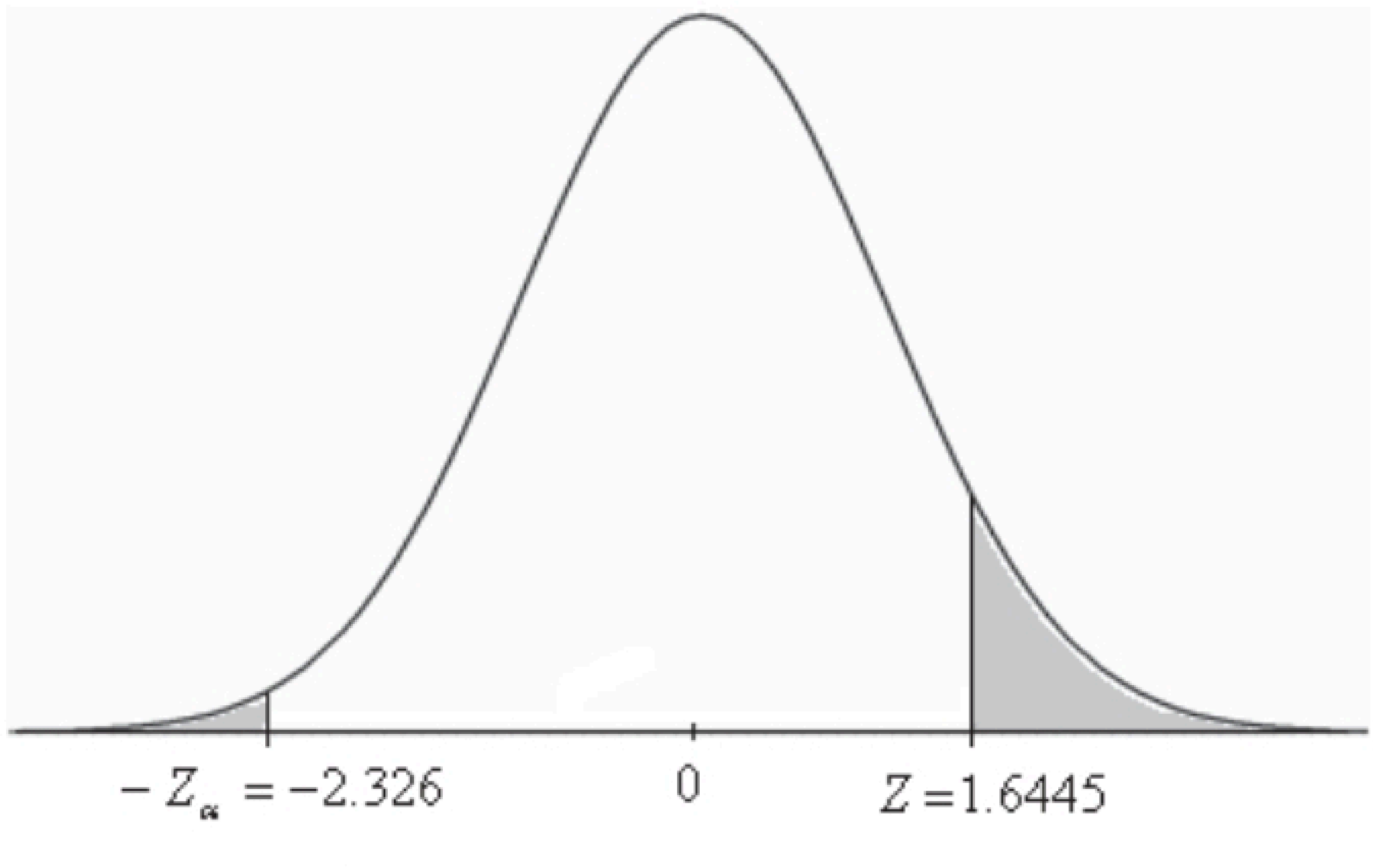
$$P(-Z_{\alpha/2} \cdot \sigma_{\overline{X}} < (\overline{X} - \mu) < Z_{\alpha/2} \cdot \sigma_{\overline{X}}) = (1-\alpha)$$

بإضافة $(-\overline{X})$ للمعادلة آنفاً نحصل على :

$$P(-\overline{X} - Z_{\alpha/2} \cdot \sigma_{\overline{X}} < -\mu < -\overline{X} + Z_{\alpha/2} \cdot \sigma_{\overline{X}}) = (1-\alpha)$$

وبالضرب في (−١) نحصل على المعادلة التاليه والتي تمثل فتره الثقه لمتوسط المجتمع μ :

$$P(\overline{X} + Z_{\alpha/2} \cdot \sigma_{\overline{X}} > \mu > \overline{X} - Z_{\alpha/2} \cdot \sigma_{\overline{X}}) = (1-\alpha) \qquad (6.7)$$



الشكل رقم (٦.١). توزيع المعاينة للمتوسط.

ولإيضاح كيفية استخدام المعادلة رقم (6.7) ، نفترض أن مجتمع إنتاجية الذرة لمزارع منطقه ما يتوزع طبيعياً بانحراف معياري يساوي (٦,٢) رطل ؛ وتم أخذ عينة عشوائية حجمها n يساوي ٣٦ مزرعة وحسبنا متوسط الإنتاجية $(\overline{X})$ والذي يساوي ١٣٢ رطل. المطلوب إيجاد فترة ثقة بنسبة ٩٥٪ لمتوسط المجتمع μ ، المتوسط الحقيقي

لإنتاجية الذرة للمنطقة.

نظراً لأن درجة الثقة المطلوبة بنسبه ٩٥٪ فإن قيم Z المناسبة هي $Z_{\alpha/2} = 1.96$ و $-Z_{\alpha/2} = -1.96$ ولذلك يمكن حساب :

$$\sigma_{\overline{X}} = \sigma / \sqrt{n} = \frac{6.2}{\sqrt{36}} = \frac{6.2}{6} = 1.033$$

وبالتعويض عن القيم في المعادله رقم (6.7) نحصل على :

$$P(\overline{X} + Z_{\alpha/2} \cdot \sigma_{\overline{X}} > \mu > \overline{X} - Z_{\alpha/2} \cdot \sigma_{\overline{X}}) = (1 - \alpha)$$

$$P(132 + (1.96)(1.033) > \mu > 132 - (1.96)(1.033) = 0.95$$

أو $$P(132 + 2.03) > \mu > 132 - 2.03) = 0.95$$

وبالتقريب نحصل على :

$$134 > \mu > 130$$

وعليه فإن متوسط إنتاجيه الذرة بين ١٣٠ و ١٣٤ رطلاً لأننا إذا كررنا سحب عينات بحجم ٣٦ من المجتمع وحسبنا فتره الثقة لمتوسط كل عينة، فإن ٩٥ عينة من أصل ١٠٠عينة سوف تعطي متوسط قريب من متوسط المجتمع μ ، وعليه فإن الفترات تشتمل على متوسط المجتمع. وليس لدينا دلالة لاعتقاد ان العينة التي تم اختيارها تختلف عن الـ ٩٥ عينة. وبالطبع فقد نكون غير محظوظين وتكون العينة المختارة ليست من الـ ٩٥ ولكن احتمال ذلك منخفض جداً. ومع ذلك، فقد تمت الإشارة للاحتمال

المصاحب لفترة الثقة حتى تكون واضحة.

ولفترة الثقة التي تم حسابها في المثال السابق كان احتمال نسبة الخطأ بافتراض أن متوسط المجتمع μ مشمول بالفتره هو α تساوي ٥٪ . وعند الرغبة في تقليل المخاطرة وتخفيض نسبة الخطأ ، فإنه يجب استخدام فتره ثقه أوسع. فمثلاً ، إذا رغبنا أن تكون نسبة الخطأ تساوي ٠,٠١ ، أو تكون فتره الثقة بنسبه ٩٩٪ ، فإن قيم Z في هذه الحالة تختلف عن السابقه ويمكن إيجادها من جدول (٧) في الملحق حيث $Z_{\alpha/2}$ تساوي ٢,٥٧ وقيمة $-Z_{\alpha/2}$ تساوي -٢,٥٧ (لـ $(\alpha / 2 = 0.005$. وباستخدام معادله (٦,٧) فإن فتره ثقه ٩٩٪ هي.

$$P(\overline{X} + Z_{\alpha/2} \cdot \sigma_{\overline{X}} > \mu > \overline{X} - Z_{\alpha/2} \cdot \sigma_{\overline{X}}) = (1 - \alpha)$$

$$132 + (2.57)(1.033) > \mu > 132 - (2.57)(1.033)$$

$$132 + 2.65 > \mu > 132 - 2.65$$

$$134.6 > \mu > 129.4$$

ويلاحظ أن فترة الثقة هذه أوسع من فترة الثقة السابقة وبذلك فإن لدينا ثقة أكبر باحتواء الفترة على المعلمة μ .

وفي حالة المجتمعات غير الطبيعيه non-normal population والتي لها انحراف معياري σ معلوم وحجم صغير n (أقل من ٣٠) فإن توزيع $Z = (\overline{X} - \mu) / \sigma_{\overline{X}}$ غير طبيعي ولذلك لا توجد لدينا طريقه مناسبة لإيجاد فترة الثقة. ولكن ، إذا كان حجم المجتمع n (أكبر من أو يساوي ٣٠) فإن نظريه النهايه المركزيه تشير إلى أن توزيع المعاينة لـ Z يتبع تقريبا التوزيع الطبيعي. وبذلك يمكن استخدام صيغة فترة الثقة التي تم استخدامها سابقاً.

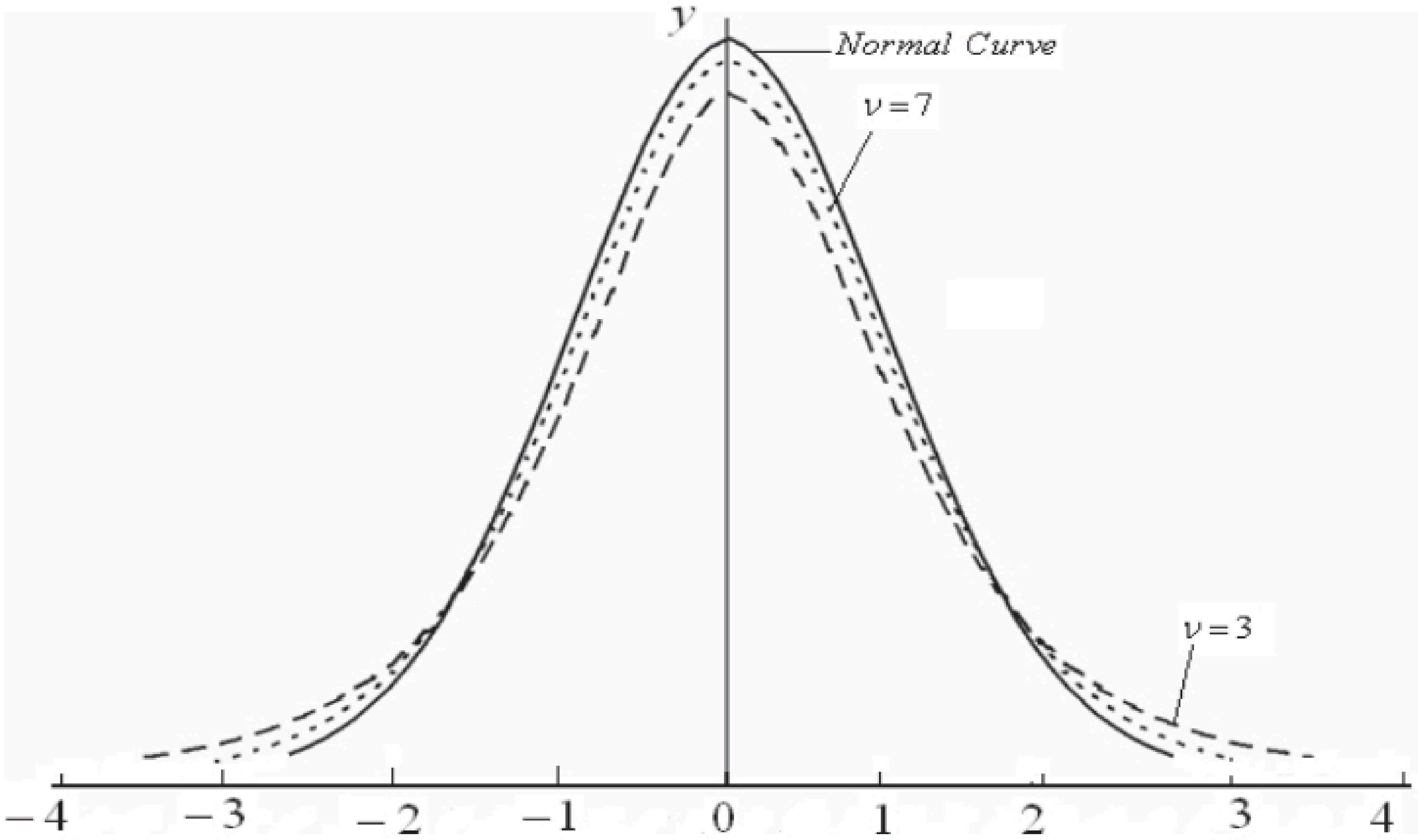
فترات الثقة لمتوسط مجتمع μ وانحرافه المعياري σ مجهول

Confidence Intervals for μ with Unknown σ

في النص السابق افترضنا معلومية الانحراف المعياري للمجتمع وذلك لتقدير متوسط المجتمع باستخدام بيانات العينة. ولكن في الغالب إذا كان الهدف هو تقدير المتوسط فإنه يجب أيضاً تقدير الانحراف المعياري من بيانات العينة. ووفقاً لهذه الظروف، فإننا نستخدم الانحراف المعياري للعينة S كتقريب للانحراف المعياري للمجتمع σ، ولذلك يتم استخدام توزيع t والموضح بالمعادلة رقم (6.8) :

$$t = (\overline{X} - \mu) / (S / \sqrt{n}) \tag{6.8}$$

بدرجة حرية $v = (n - 1)$ (وبافتراض أن المجتمع يتبع التوزيع الطبيعي) وتوزيع t عبارة عن عائلة من دوال الكثافة الاحتمالية والتي يعتمد شكلها على درجة الحريه v (v عبارة عن حرف أغريقي يستخدم للتعبير عن درجة الحرية) وليس على الانحراف المعياري للمجتمع σ. ويتميز توزيع t بأن له شكل يشبه الجرس مقارب للتوزيع الطبيعي ولكنه ليس مدبب تماماً مثله، ولذلك فهو أكثر اتساعاً في الأطراف (الشكل رقم ٦،٢) .



الشكل رقم (٦.٢). توزيع t لقيم مختارة لـــدرجات الحرية.

ونظراً لأن درجة الحرية ν تحدد شكل توزيع t وهي تزيد بزيادة حجم العينة، فإن توزيع t يصبح أكثر تدبباً كلما زاد حجم العينة. وعندما تؤول قيمة درجات الحرية إلى ما لانهاية فإن توزيع t يصبح متقارباً مع التوزيع الطبيعي. وفي الحقيقة فإن قيم t (من الملحق جدول رقم ٨) عندما تكون درجة الحرية مالا نهاية تكون هي نفسها قيمه z. أما في حالة العينات الصغيرة فإن قيم t تكون أكبر من قيم z للتوزيع الطبيعي لنفس المستوى من المعنويه α. والسؤال الآن متى يكون التوزيعين متقاربين نسبياً بحيث يكون الخطأ الناتج من استخدام التوزيع الطبيعي كتقريب لتوزيع t غير حرج ؟ والإجابة التي قررها معظم الإحصائين هي عندما يكون حجم العينه n مساوٍ لـ ٣٠. وعليه، فإنه للعينات الصغيره $(n < 30)$ يتم استخدام توزيع t عادة كقاعدة لايجاد فترات الثقة واختبارها بالنسبة لـ μ ، وفي حالة العينات الكبيره يستخدم التوزيع الطبيعي كتقريب. للعينات الصغيرة (حجم العينه n أقل من ٣٠) يمكن الحصول على فترة الثقة لمتوسط

المجتمع μ باستخدام توزيع t كالتالي : يمثل التعبير $t_{(\alpha / 2, v)}$ قيمة توزيع t بدرجة حريه $v = n - 1$ والتي تحدد قيمة الاحتمال عند $\alpha/2$ في الطرف العلوي أي أن $P(t > t_{(\alpha / 2, v)}) = \alpha/2$ وبصفة مماثله فإن $P(t < -t_{(\alpha / 2, v)}) = \alpha/2$ للطرف السفلي ؛ ولذلك يمكن التعبير عن ذلك رياضياً بالمعادلة التالية :

$$P(-t_{(\alpha / 2,v)} < t < t_{(\alpha / 2,v)}) = (1 - \alpha) \tag{6.9}$$

وحيث إن t قد تم تعريفها في المعادلة رقم (6.9) سابقاً فإنه يمكن التعويض بقيمتها في المعادله آنفاً وحل المعادلة لإيجاد فترة الثقة لمتوسط المجتمع μ بدرجه ثقه $100\,(1 - \alpha)$ والتي يمكن التعبير عنها رياضياً كالتالي :

$$\overline{X} - t_{(\alpha / 2,v)} \cdot S/\sqrt{n} < \mu < \overline{X} + t_{(\alpha / 2, v)} \cdot S/\sqrt{n} \tag{6.10}$$

ويمكن إيضاح تطبيق الصياغة آنفاً من خلال المثال التالي :

نفترض أن أحد علماء الانتاج الحيواني يرغب في تقدير متوسط الوزن عند الفطام بعد ٢٠٥ أيام من الرضاعة لمجموعة عجول لها نفس الأب. حيث قام بسحب عينة عشوائية حجمها ٩ عجول من القطيع وتحصّل على البيانات التالية :

550, 525, 570, 600, 485, 535, 580, 520, 540

والمطلوب إيجاد فتره ثقه بدرجة ٩٥٪ لمتوسط الوزن عند الفطام.

ولعمل ذلك فإننا نقوم أولاً بحساب قيم متوسط العينه $\overline{X}$ والانحراف المعياري S ، ثم بعد ذلك نستخدم جدول رقم (٨) في الملحق للحصول على قيم $\pm\ t_{(\alpha / 2,v)}$. وأخيراً نعوض بتلك القيم في قانون فترة الثقة للحصول على النتيجة. ويوضح الجدول رقم (٦.١) الحسابات اللازمه لإيجاد المتوسط والتباين لبيانات الوزن عند الفطام.

الجدول رقم (٦.١). حساب المتوسط والتباين لبيانات الوزن المكتسب.

X (الوزن)	$X - \overline{X}$	$(X - \overline{X})^2$
٥٥٠	٥	٢٥
٥٢٥	٢٠ -	٤٠٠
٥٧٠	٢٥	٦٢٥
٦٠٠	٥٥	٣٠٢٥
٤٨٥	٦٠ -	٣٦٠٠
٥٣٥	١٠ -	١٠٠
٥٨٠	٣٥	١٢٢٥
٥٢٠	٢٥ -	٦٢٥
٥٤٠	٥ -	٢٥
٤٩٠٥	٠	٩٦٥٠

ومن البيانات آنفاً فإن :

$$\overline{X} = \sum X/n = 4905/9 = 545$$
$$S^2 = \sum(X - \overline{X})^2/n - 1 =$$
$$9650/(9-1) = 9650/8 = 1{,}206.25$$
$$S = \sqrt{S^2} = \sqrt{1{,}206.25} = 34.73$$
$$v = (n - 1) = (9 - 1) = 8$$

وحيث إن $\alpha = 0.05$ وعليه فإن $t_{(\alpha/2,v)} = t_{(0.05/2,8} = t_{(0.025,8)} = 2.306$

وبالتعويض بالقيم سابقاً في المعادله رقم (6.10) نحصل على :

$$\overline{X} - t_{(\alpha/2,v)} \cdot S/\sqrt{n} < \mu < \overline{X} + t_{(\alpha/2,v)} \cdot S/\sqrt{n}$$
$$545 - 2.306 \cdot 34.73/\sqrt{9} < \mu < 545 + 2.306 \cdot 34.73/\sqrt{9}$$
$$545 - 2.306(11.58) < \mu < 545 + 2.306(11.58)$$
$$545 - 26.7 < \mu < 545 + 26.7$$
$$518.3 < \mu < 571.7$$

الفترة من ٥١٨,٣ إلى ٥٧١,٧ تمثل فترة الثقة بدرجة ٩٥٪ لمتوسط الوزن للعجول ذات العمر ٢٠٥ يوم التي لها نفس الأب. ونظراً لأنها كبيرة نوعاً ما فإن الأمر يتطلب زيادة حجم العينة. ولكننا في الغالب لانتعامل مع العينات ذات الحجم الكبير في جميع الحالات ؛ لأننا نختار بعض العينات لإجراء الاختبارات على بيانات العينة ونحاول معرفة التكاليف المصاحبة للعينات الكبيرة.

وفي حالات أخرى ، فإن بعض العوامل مثل تكاليف الحصول على العينة أو مجموعة الخبراء اللازمين لاجراء المعاينة تجعل العينات الصغيره الحل العملي المناسب لتقدير المعالم.

للعينات الكبيرة حجم العينة أكبر من أو يساوي ٣٠ ، $(n \geq 30)$ يستخدم تقريب التوزيع الطبيعي لتوزيع t كقاعدة لتحديد فترة الثقة للمتوسط μ بدرجة ثقة $100(1-\alpha)$. وبعبارة أخرى نستخدم S (الانحراف المعياري للعينة) كمؤشر للانحراف المعياري للمجتمع σ في إيجاد فترة الثقة كما تم عرضه سابقاً أي :

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = (1-\alpha) \tag{6.11}$$

ومن الجدول آنفاً و بمعلومية :

$$z = (\overline{X} - \mu) /(S/\sqrt{n})$$

فإنه يمكن إيجاد فترة الثقة باستخدام المعادلة التالية :

$$\overline{X} - Z_{\alpha/2} \cdot (S/\sqrt{n} < \mu < \overline{X} + Z_{\alpha/2} \cdot (S/\sqrt{n}) \tag{6.12}$$

وبافتراض أننا نرغب في زيادة حجم العينه في مثال العجول السابق فقد تمت زيادة حجم العينه من ٩ عجول إلى ٣٦ عجلاً وحسبنا المتوسط فوجد أنه يساوي $\overline{X} = 548$

والانحـراف المعيـاري $S = 20.5$ رطـل ، وبـذلك يمكـن اسـتخدام المعادلـة رقـم (6.12) لإيجاد فتره الثقه لمتوسط المجتمع μ بدرجه ثقة ٩٥٪ كالتالي :

$$\overline{X} - Z_{\alpha/2} \cdot (S/\sqrt{n} < \mu < \overline{X} + Z_{\alpha/2} \cdot (S/\sqrt{n})$$

$$548 - 1.96 \cdot (20.5/\sqrt{36}) < \mu < 548 + 1.96 \cdot (20.5/\sqrt{36})$$

$$548 - 1.96(3.42) < \mu < 548 + 1.96(3.42)$$

$$548 - 6.7 < \mu < 548 + 6.7$$

$$541.3 < \mu < 554.7$$

ويلاحظ من النتائج أن الفرق بين الحد الأعلى والحد الأدنى للثقة بعد زيادة حجـم العينـه يساوي ١٣,٤ رطل وهو أقل بكثير مـن الفـرق في حالـة العينـة الـصغيرة. وقـد حـدث هـذا نتيجة لمجموعة من العوامل أولها أنه في حالـة العينـات الـكبيره فإننـا نـستخدم قـيم جـدول التوزيع الطبيعي القياسي لايجاد قيم $Z_{\alpha/2}$ والتي نوعاً ما أقل من قيم t . والعامل الثاني أنه في حالة العينات الكبيرة فإننا نحصل على قيم أقل للانحراف المعياري للعينه S والذي يحصل عادة. وأخيراً فإن العينات الكبيره ينتج عنها قيم أقل لـ $S_{\overline{X}}$ ؛ لأننا نقسم الانحراف المعياري للعينة على الجذر التربيعي للعدد ٣٦ بدلاً من قسمته على الجذر التربيعي للعدد ٩ في حالة العينات الصغيره.

فترات الثقة لــ P من التوزيع ثنائي الحدين

Confidence Intervals for *p* from The Binomial Distribution

المتغير العـشوائي ثنـائي الحـدين r يعبّـر عـن عـدد مـرات النجـاح في عـدد n مـن المحـاولات للتجربـه ، وكمـا هـو معلـوم مـن الفـصول الـسابقة فـإن متوسـط وتبـاين ثنـائي الحدين يعبّر عنه بالمعادله التالية :

$$E(r) = np \text{ ، } Var(r) = npq, \qquad (6.13)$$

في بعض الأحيان قد يكون الاهتمام منصب على متغير آخر، $\hat{p}$ ، والذي يمثل نسبة النجاح في عدد n من المحاولات حيث تعبر n عن العينة المختارة من مجتمع يتبع للتوزيع ثنائي الحدين. ويتم الحصول على $\hat{p}$ بقسمة r على n ، كما هو موضح في المعادلة التالية:

$$\hat{p} = (1/n).r \quad \text{أو} \quad \hat{p} = r/n \qquad (6.14)$$

وحيث إن (1/n) قيمة ثابتة فإن القيمة المتوقعة لـ $\hat{p}$ تساوي:

$$E(\hat{p}) = (1/n).E(r) = (1/n).np = p$$

وكذلك التباين يمكن التعبير عنه بالمعادلة التالية:

$$Var(\hat{p}) = (1/n)^2.Var(r) = (1/n)^2.npq = pq/n$$

وعليه فإن $\hat{p}$ عبارة عن مقدّر لنسبة النجاح في المجتمع، p. ويمكننا ببساطة إيجاد فترة ثقة بنسبة $100(1-\alpha)$ وذلك باستخدام تقريب التوزيع الطبيعي لثنائي الحدين. ويتم ذلك باعتبار أن متوسط التوزيع ثنائي الحدين مساوٍ لمتوسط التوزيع الطبيعي وأن الانحراف المعياري لثنائي الحدين مساوٍ للانحراف المعياري للتوزيع الطبيعي واستخدام صياغة التوزيع Z. وبناء على ذلك فإن $E(\hat{p}) = p$ و الانحراف المعياري لـ $\hat{p}$ يساوي $\sqrt{pq/n}$. ونظراً لوجود مقدر مجهول p نهدف لتقديره فإنه يتعذر استخدام توزيع Z لذلك نستخدم $\hat{p}$ بدلاً من p و $(1-\hat{p})$ بدلاً من q ليصبح الانحراف المعياري $\sqrt{[(\hat{p})(1-\hat{p})/n]}$ وتصبح معادلة توزيع z على النحو التالي:

$$Z = (\hat{p} - p)/\sqrt{[(\hat{p})(1-\hat{p})/n]} \qquad (6.15)$$

وبالتعويض بالمعادلة السابقة في المعادلة رقم (6.11) وإعادة ترتيب الحدود نحصل على فترة الثقة للمقدر p والتي يمكن التعبير عنها رياضياً كالتالي :

$$\hat{p} - Z_{\alpha/2}.\sqrt{[(\hat{p})(1-\hat{p})/n]} < p < \hat{p} + Z_{\alpha/2}.\sqrt{[(\hat{p})(1-\hat{p})/n]}. \qquad (6.16)$$

ولتوضيح ذلك نفترض أن أحد علماء الحيوان والذي قام باختيار عينة حجمها ١٠٠ بقرة من قطيع كبير تم تلقيحه صناعياً. وخلال الاختبار حدد أن ٥٩ تم تلقيحها ومن ثم حملها. المطلوب إيجاد فترة ثقة ٩٥٪ لـ p والتي تعبر عن نسبة الأبقار في القطيع التي حملت. ولحل مثل هذه المسألة فإننا نوجد قيمة $Z_{\alpha/2}$ الجدولية باستخدام الجدول رقم (٧) في الملحق حيث نجد أن قيمة $Z_{0.025}$ تساوي ١,٩٦ وبحساب قيمة $\hat{p} = r/n$ نجد أنها تساوي ٥٩/١٠٠=٠,٥٩ ثم التعويض بتلك القيم في المعادلة رقم (6.16) كالتالي :

$$\hat{p} - Z_{\alpha/2}\sqrt{[(\hat{p})(1-\hat{p}/n]} < p < \hat{p} + Z_{\alpha/2}.\sqrt{[(\hat{p})(1-\hat{p})/n]}$$
$$0.59 - 1.96.\sqrt{[(0.59)(1-0.59)/100]} < p < 0.59 + 1.96.\sqrt{[(0.59)(1-0.59/100]}$$
$$0.59 - 1.96.\sqrt{0.002419} < p < 0.59 + 1.96.\sqrt{0.002419}$$
$$0.59 - 1.96 .0.049 < p < 0.59 + 1.96 .0.049$$
$$0.59 - 0.096 < p < 0.59 + 0.096$$
$$0.494 < p < 0.686$$

ولذلك تقدير النسبة الفعلية للأبقار التي تم حملها تتراوح ضمن المدى من ٠,٤٩ إلى ٠,٦٩ للقطيع وذلك بعد القيام بتلقيح القطيع صناعياً وبدرجة ثقة ٩٥٪.

فترة الثقة للتباين σ^2

Confidence Interval for σ^2

في بعض الحالات، قد نحتاج لإيجاد تقدير لتباين مجتمع غير معلوم. فعلى سبيل المثال، الأدوات العلمية يجب أن تعطي قراءات غير متحيزة بأخطاء قياس قليلة جداً. فمثلاً منظار التربة والذي يعطي متوسط قراءات صحيحة يقل استخدامه إذا كان الاختلاف في القراءات كبيراً وبذلك انعدام الاتساق من عينة لعينة من نفس التربة. وكذلك في مصانع الآلات الزراعية، لا يقتصر الأمر على صحة متوسط حجم الأجزاء المنتجة بل يجب أن يكون الاختلاف في الحجم قليل ويمكن التحكم به أو لن تكون مطابقة للمقاييس وعليه تكون الأجزاء غير قابلة للبيع. من جهة أخرى فإن الاقتصاديين الزراعيين يهتمون بتقلبات الأسعار عند إعداد الخطط التسويقية للخضروات والفواكه للمجموعات المنتجة. وتباين العينة والممكن حسابه باستخدام المعادلة رقم (6.17)، عبارة عن مقدّر غير متحيز لتباين المجتمع σ^2 .

$$\mathrm{S}^2 = \frac{\sum(X-\overline{X})^2}{\mathrm{n}-1} \tag{6.17}$$

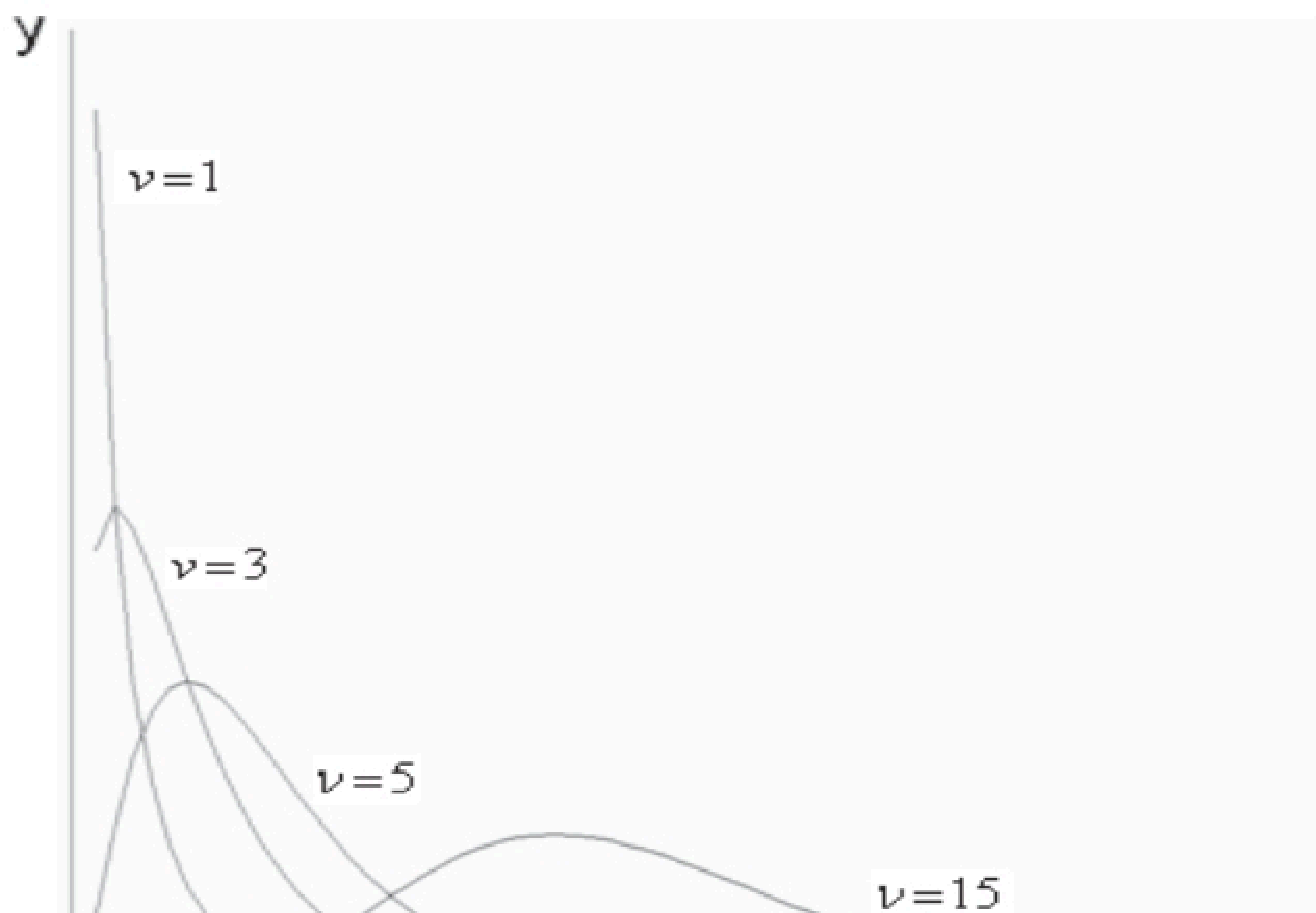
ويتم الحصول على تباين العينة باستخدام عينات تنتمي لدوال الكثافة الاحتمالية pdfs لمربع كاي. وقد تم قياس (معايرة) تباين العينة S^2 بطريقة مشابهة للمتغير z. وينتمي المتغير العشوائي القياسي χ^2 والمعرَّف بالمعادلة رقم (٦.١٨) لتوزيع مربع كاي إذا كان مجتمع المتغير العشوائي x والذي حسب منه تباين العينة S^2 يتوزع طبيعياً.

$$\chi^2 = \frac{(\mathrm{n}-1)\,\mathrm{S}^2}{\sigma^2} \tag{6.18}$$

ويوضح الجدول رقم (٩) بالملاحق احتمالات الأطراف لتوزيع مربع كاي. ويختلف توزيع تباين العينة S^2 عن توزيع المتوسط $\overline{X}$ بأنه غير متماثل؛ نظراً لأن S^2

لا يمكن أن يأخذ قيمة سالبة. تم فصل احتمالات كل طرف من توزيع مربع كاي. وقد تم التعبير عن قيم احتمالات الطرف الأيمن بالرمز $\chi^2{}_{\alpha}$ و قيم احتمالات الطرف الأيسر بالرمز $\chi^2{}_{1-\alpha}$ لهذا التوزيع.

ويعتمد شكل هذا التوزيع ، مثل توزيع t، على درجات الحرية v المرتبطة بالتباين S^2. لذا فإن تغيير حجم العينة سيؤدي لتغير درجات الحرية وعليه تغير دوال الكثافة الاحتمالية لمربع كاي (الشكل رقم ٦٫٣). ومن ثمَّ كلما زاد حجم العينة فإن شكل توزيع مربع كاي يصبح مقارب للتوزيع الطبيعي. ولذلك عندما تكون درجة الحرية كبيرة، مثلاً ١٠٠فإن مربع كاي يبدو مقارب جداً لشكل التوزيع الطبيعي مقارنة بدرجات الحرية الصغيرة ،٦ درجات مثلاً.



الشكل رقم (٦.٣). توزيع مربع كاي لمجموعة مختارة من درجات الحرية.

ولإيجاد فترة الثقة بنسبة $100(1-\alpha)$ لتباين المجتمع σ^2 عندما تكون العينة من مجتمع طبيعي يتم تكوين التعبير الموضح في المعادلة رقم (6.19).

$$P(\chi^2_{1-\alpha/2,v} < \chi^2 < \chi^2_{(\alpha/2,v)}) = (1-\alpha) \qquad (6.19)$$

ثم نستخدم تعبير مربع كاي الموضح في المعادلة رقم (6.18) ونعوض به في التعبير الاحتمالي ، معادلة رقم (6.20)، ونرتب الحدود لنحصل على فترة الثقة المطلوبة معادلة رقم (6.21).

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

$$\chi^2_{(1-\alpha/2,v)} < (n-1)S^2/\sigma^2 < \chi^2_{(\alpha/2,v)} \qquad (6.20)$$

$$\frac{(n-1)S^2}{\chi^2_{(\alpha/2,v)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2,v)}} \qquad (6.21)$$

ولإيجاد المعادلة رقم (6.21)، تم تغيير مفهوم اللامساوة بحيث أصبحت قيم مربع كاي الكبيرة الجدولية جزء من طرف المعادلة الأيسر وقيم الجدول الصغيرة جزء من طرف المعادلة الأيمن. ويمكن توضيح استخدام تلك المفاهيم من خلال المثال التالي :

ترغب إحدى تعاونيات الألبان بشراء آلة لتعبئة و تغليف الحليب في عبوات سعة نصف جالون. وقام مندوب مبيعات شركه Sur-Fil بعرض للمنتج وأفاد بأن مدى الاختلاف بين العبوات لايتجاوز ٠,٤ أوقيه وللتحقق من ذلك قام المدير المسئول بأخذ عينه من خط الإنتاج حجمها ٨ عبوات سعة نصف جالون وحسب المتوسط والتباين للعينه فوجدهما ٦٤,١ و ٠,٠١٨ على التوالي. المطلوب إيجاد فتره ثقه ٩٠٪ للتباين σ^2.

الحل :

نبــدأ أولاً بإيجــاد درجــة الحريــه والــتي يمكــن حســـابها كالتـــالي $v = (n - 1) = (8 - 1) = 7$. ومن الجدول رقم (٩) في الملحق نجد أن قيم $\chi^2_{0.05}$ و $\chi^2_{0.95}$ عند درجه حرية ٧ هي ١٤,٠٧ و ٢,١٧ على التوالي.

وباستخدام المعادلة رقم (6.21) يمكن ايجاد فتره الثقه كالتالي :

$$\frac{(n-1)S^2}{\chi^2_{(\alpha/2,v)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2,v)}}$$

$$(7)(0.018)/14.07 < \sigma^2 < (7)(0.018)/2.17$$
$$0.009 < \sigma^2 < 0.058$$

ويمكن تفسير ذلك بأن التباين بين العبوات يبدو صغيراً ولكن إذا نظرنا للحد الأعلى للعينة وأخذنا الجذر التربيعي لها للحصول على الانحراف المعياري نجد أنه يساوي ٠,٢٤ أوقية ولذلك فعند إنشاء مدى يساوي ضعف الانحراف المعياري فإن المدى المذكور بواسطة مندوب المبيعات يصبح محل تساؤل. وعليه فقبل اتخاذ قرار الشراء ربما يتطلب الأمر حصول المدير على معلومات إضافية.

تمارين Exercises

١- سحبت عينة عشوائية بسيطة حجمها ٤٠٠ من مزارع القطن فوجد أن متوسط المساحة يساوي ٦١٥ بانحراف معياري يساوي ١٠٠. أوجد فترة ثقة بنسبة٩٠٪ لمتوسط المجتمع ثم فسّرها.

٢- في محاولة لدراسة مدى مطابقة أصواف الأغنام والماعز للمواصفات الجديدة الخاصة بالصوف سحبت عينة عشوائية بسيطة حجمها ١٠٠٠ من مربين للماعز والأغنام فوجد أن ٦٠٠ منها مطابقة للمواصفات في حين أن ٤٠٠ لم تطابق المواصفات. اوجد فترة الثقة بنسبة ٩٥٪ للنسبة الحقيقية لجميع مربي الأغنام

والماعز وفقاً للمواصفات الجديـــدة.

٣- قام مصنع الآلات الزراعية بإعداد طلبيه كبيرة من مسامير المحاريث. وعند وصولها أخذ قسم التحكم بالجودة بالمصنع عينة عشوائية منها حجمها ٣٠٠ مسمار فوجد أن ٣٦ منها معيبة. أوجد فترة ثقة بنسبة ٩٩٪ لنسبة المسامير المعيبة في الشحنة.

٤- وجد أن متوسط محيط الخصيتين للثيران الرضيعة لعينة عشوائية حجمها ١٠٠ مأخوذة من قطيع الأبقار يساوي ٣١,٥سم بانحراف معياري ١٠ سم. أوجد فترة الثقة بنسبة ٩٥,٥٪ لمتوسط محيط الخصيتين لتلك الثيران.

٥- في دراسة لتقييم المني للثيران الرضيعة لقطيع معين أخذت عينة حجمها ٩ فوجد أن متوسط النسبة للحيوانات المنوية الأولية الغير طبيعية يساوي ١٨ بانحراف معياري ٢,٧. أوجد فترة الثقة لمتوسط النسبة الحقيقية للحيوانات المنوية الأولية غير الطبيعية في هذا القطيع.

٦- أخذت عينة من نباتات القطن لبعض الأصناف الخاصة عددها ٢٥ لدراسة الإثمار. وجد أن متوسط عدد اللوزات الناضجة للنبات يساوي ٢٠ بانحراف معياري ١٠. أوجد فترة الثقة بدرجة ٩٩٪ لعدد اللوزات الناضجة الحقيقي للنبات لهذا الصنف.

٧- في دراسة لإنبات نباتات مستوردة من المهم المحافظة على مستوى الرطوبة النسبية في الحدود التي يتحملها النبات لتجنب فشل عملية الإنبات. قام البستاني بأخذ عدد ٥٠ قراءة ووجد أن التباين في الرطوبة النسبية يساوي ٠,٦ ٪ . أوجد فترة الثقة بنسبة ٩٩٪ للاختلاف الحقيقي للرطوبة داخل البيت المحمي.

٨- قام ستة محكمين بتقييم مجموعة حظائر للثروة الحيوانية من حيث شكلها الحديث. ووجد أن التباين في مستوى التحكيم يساوي ١٠. أوجد فترة الثقة بنسبة ٩٥٪ للاختلاف الحقيقي في مستوى التحكيم.

الفصل السابع

التقدير الإحصائي لعينتين

Statistical Estimation: Two Samples

في بعض الأحيان نحتاج لمقارنة منتجين أو طريقتين لتقييم المهمات أو الخدمات، وعليه فإنه يجب علينا جمع بيانات من عينتين مختلفتين والتعامل مع متغيرين عشوائيين. ولكنّه لا يمكننا تقدير تلك المتغيرين؛ نظراً لعدم وجود توزيعات احتماليه يمكن استخدامها في هذه الحالة. لذلك يمكن إجراء تعديل بسيط في المتغيرات مما يمكننا من استخدام التوزيعات الاحتمالية المشار لها في الفصل السادس. فعلى سبيل المثال، بدلاً من التعامل مع متغيرين عشوائيين $\overline{X}_1$ ، $\overline{X}_2$ واستخدامها بصورة منفصلة لتقدير متوسط المجتمعتين موضوع الدراسة يمكن تحويلها لمتغير عشوائي واحد بطرح أحدهما من الآخر. ويصبح المتغير العشوائي الجديد هو $(\overline{X}_1 - \overline{X}_2)$ والذي ينتمي لتوزيع احتمالي طبيعي ذو متغير واحد متوسطه $(\mu_1 - \mu_2)$ وتباينه $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$. والاختلاف الوحيد في هذه الحالة هو التباين حيث إنه لا يمكن أن يكون سالباً وعليه تم جمع تباين المتغيرين العشوائيين.

ويمكن ببساطة الحصول على تقديرات النقطة بحساب قيم المتغيرات العشوائية لكل عينة. وسيتم عرض فترات الثقة في الأجزاء القادمة لمجموعة من المتغيرات العشوائية لكل منها توزيع معاينة مختلف.

فترات الثقة للفرق بين متوسطي مجتمعين $(\mu_1 - \mu_2)$ لها انحرافات معيارية معلومة

Confidence Intervals for $(\mu_1 - \mu_2)$ with σ_1 σ_2 Known

نفترض أن المنتج يرغب في مقارنة متوسط عدد أيام الشبق لمجموعتين مختلفة من منتجات الهرمونات المعطاة لقطيع الماشية. قام المنتج بأخذ عينة عشوائية من المجتمع الأول والذي يساوي متوسطة μ_1 وانحرافه المعياري σ_1 حيث تمت معاملة الحيوانات منها بالجرعة الأولى وحسب متوسط العينة $\overline{X}_1$ ، ثم أخذت عينة عشوائية ثانية مستقلة من المجتمع الثاني والذي متوسطه μ_2 وانحرافه المعياري σ_2 حيث تمت معاملة الحيوانات في هذا المجتمع بالجرعة الثانية ثم حسب متوسط العينة $\overline{X}_2$.

فإذا قام المنتج بأخذ عينات متكررة من المجموعين بحجم n_1 ، n_2 وقام بحساب المقدر $(\overline{X}_1 - \overline{X}_2)$ لكل تجربة فإنه سيحصل على توزيع معاينة طبيعي بمتوسط يساوي الفرق بين متوسطات المجتمع وتباين يساوي مجموع تباين المجتمعين. وعليه فإنه يمكن حساب فترة الثقة للفرق بين متوسطي المجتمع $(\mu_1 - \mu_2)$ بدرجة ثقة تساوي $100(1-\alpha)$:

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = (1-\alpha) \tag{7.1}$$

وبالتعويض في المعادلة أعلاه بقيمه z والتي تساوي:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_2 - \mu_2)}{\sqrt{\frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}}} \tag{7.2}$$

حيث المقام يعبّر عن الخطأ المعياري للفرق بين المتوسطات ويرمز له بالرمز $\sigma_{(\bar{x}_1 - \bar{x}_2)}$. وهذا الحد يدخل ضمن المعادلة التي نحسب منها فترة الثقة بدرجة ثقة

$100\,(1-\alpha)$ والتي يمكن التعبير عنها بالمعادلة التالية :

$$(\overline{X}_1-\overline{X}_2)-Z_{\alpha/2}.\sigma(\bar{x}_1-\bar{x}_2)<(\mu_1-\mu_2)<(\overline{X}_1-\overline{X}_2)+Z_{\alpha/2}.(\sigma_{\bar{x}_1-\bar{x}_2}) \quad (7.3)$$

والتي تمثل فتره الثقة للفرق بين متوسطات المجتمع عندما يكون الانحراف المعياري لمجتمع الدراسة معلوم. ويمكن إيضاح ذلك من خلال المثال التالي :

قامت إحدى المختصات بدراسة التربة بإجراء اختبار معملي لمحتويات العنب الأحمر من السكر فوجدت أن الانحراف المعياري له يساوي خمسة $(\sigma_1=5)$ عند زراعته في تربة رملية بينما يكون الانحراف المعياري له يساوي ٧,٧٥ $(\sigma_2=7.75)$ عند زراعته في تربه رمليه طينية. وحيث إن الهدف من ذلك هو تقدير متوسط الفرق في محتويات السكر لأنواع التربة ؛ بهدف تحديد شروط الزراعة بشكل تجاري. لذا ، قامت الباحثة بأخذ عينة عشوائية من حقول العنب المزروعة في التربة الرملية حجمها ٢٥ $(n_1=25)$ فوجدت أن متوسط السكر فيها يساوي ٢١ $(\overline{X}_1=21)$ وأخذت عينه عشوائية أخرى من حقول العنب المزروع في التربة الرملية الطينية حجمها ٣٠ $(n_2=30)$ فوجدت أن متوسط السكر فيها يساوي ٢٦ $(\overline{X}_2=26)$. والمطلوب حساب فتره الثقة بدرجه ٩٨٪ لفرق المتوسط للسكر في العنب الأحمر المنتج بشكل تجاري في هذه التربة.

لكتابة فتره الثقة يجب أولاً أن نحسب قيمة $Z_{\alpha/2}$ من الملحق جدول (٧)، والتي تساوي ±٢,٣٣ ثم نحسب الخطأ المعياري للفرق في المتوسطات كالتالي :

$$\sigma_{(\bar{x}_1-\bar{x}_2)}=\sqrt{\sigma_1{}^2/\mathrm{n}_1+\sigma_2{}^2/\mathrm{n}_2)}=\sqrt{(5^2/25+7.75^2/30)}=\sqrt{3.002}=1.73$$

بعد ذلك يتم التعويض في المعادلة رقم (7.3) بالقيم المتحصل عليها آنفاً لنحصل على فتره الثقة بدرجة ٩٨٪ والتي تساوي :

$$(\overline{X}_1 - \overline{X}_2) - Z_{\alpha/2} \cdot \sigma_{(\bar{x}_1 - \bar{x}_2)} < (\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + Z_{\alpha/2} \cdot \sigma_{(\bar{x}_1 - \bar{x}_2)}$$
$$(21 - 26) - 2.33 \cdot (1.73) < (\mu_1 - \mu_2) < (21 - 26) + 2.33 \cdot (1.73)$$
$$-5 - 4.03 < (\mu_1 - \mu_2) < -5 + 4.03$$
$$-9.03 < (\mu_1 - \mu_2) < -0.97.$$

وعليه، فإنه يمكن القول وبدرجه ثقة ٩٨٪ إن متوسط السكر للعنب الأحمر المنتج تجارياً في التربة الرملية الطينية أكبر بمعدل يتراوح بين ٠,٩٧ و ٩,٠٣ وحدة عند مقارنته بما يتم إنتاجه في التربة الرملية.

فترات الثقة للفرق بين متوسطي مجتمعين $(\mu_1 - \mu_2)$ لها انحرافات معيارية غير معلومة

Confidence Intervals for $(\mu_1 - \mu_2)$ with σ_1 σ_2 Unknown

لتقدير الفرق بين متوسطي مجتمعين عندما يكون الانحراف المعياري لهما غير معلوم نستخدم توزيع t ولكن في الحقيقة يعتمد استخدامه على حجم العينات. فإذا كانت العينتان كبيرتين، أي أن حجمهما أكبر من ٣٠ $(n_1 , n_2 > 30)$، فإنه يمكن استخدام تباين العينات كمقدّر لتباين المجتمع، في حين أنه لا يمكن عمل ذلك في حالة العينات الصغيرة.

العينات الكبيرة Large Samples

في العينات الكبيرة يستخدم التوزيع الطبيعي كتقريب لتوزيع t لتقدير الفرق بين المتوسطات ويمكن إيجاد فتره الثقة بطريقة مشابهة تقريباً للجزء السابق ماعدا أنه يتم حساب Z باستخدام الصيغة الرياضية الموضحة في المعادلة رقم (7.4) :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2_1}{n_1} + \frac{S^2_2}{n_2}}} \tag{7.4}$$

حيث يتم التعويض بها في المعادلة التالية :

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = (1 - \alpha)$$

ومن ثم حلّها رياضياً للحصول على فتره الثقة المطلوبة والموضحة في المعادلة رقم (7.5) :

$$(\overline{X}_1 - \overline{X}_2) - Z_{\alpha/2} \cdot S_{(\overline{x}_1 - \overline{x}_2)} < (\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + Z_{\alpha/2} \cdot S_{(\overline{x}_1 - \overline{x}_2)} \tag{7.5}$$

حيث إن $S_{(\overline{X}_1 - \overline{X}_2)} = \sqrt{S_1^2/n_1 + S_2^2/n_2}$ عبارة عن تقدير للخطأ المعياري للمجتمع للفرق بين المتوسطات بناء على تباين العينة. ويمكن عرض ذلك من خلال المثال التالي :

يمكن صناعة جزء من المحراث باستخدام عمليات البثق أو يمكن صنعه باستخدام آلة القوالب ثم تلحيمه ويرغب مدير المشروع في معرفة قوة ذلك الجزء المصنوع بالطريقتين حيث تتوفر لديه مجموعة من الأجزاء مصنوعة وقام بحساب الإحصاءات الموضحة في الجدول رقم (٧,١). والمطلوب إيجاد فتره الثقة بدرجة ثقة ٩٩٪ لمتوسط الفرق لقوة الجزء المصنوع بكلا الطريقتين.

ولإيجاد فتره الثقة المطلوبة فإننا أولاً نوجد $Z_{\alpha/2}$ من الجدول رقم (٧) بالملحق عند مستوى معنوية $\alpha = 0.01$ حيث إن القيمة تساوي $\pm$ ٢,٥٧.

وبالتعويض في معادلة الانحراف المعياري بالقيم المذكورة أعلاه نستطيع حساب قيمة الانحراف المعياري كالتالي :

$$S_{(\bar{x}_1-\bar{x}_2)} = \sqrt{S_1^{\,2}/n_1 + S_2^{\,2}/n_2} = \sqrt{(256/64+171.5/49)} = \sqrt{7.5} = 2.74$$

الجدول رقم (٧.١). بيانات العينة المتحصل عليها لطريقي البثق و آله القوالب لصناعة جزء المحراث.

البند	طريقه البثق	طريقه الطبع والتلحيم
حجم العينة	٦٤	٤٩
متوسط القوة	٥١٠	٤٩٠
تباين العينة	٢٥٦	١٧١.٥

ومن ثم يتم التعويض بالقيم في المعادلة رقم (7.5) كالتالي :

$$(\bar{X}_1-\bar{X}_2) - Z_{\alpha/2}.S_{\bar{x}_1-\bar{x}_2} < (\mu_1-\mu_2) < (\bar{X}_1-\bar{X}_2) + Z_{\alpha/2}.S_{\bar{x}_1-\bar{x}_2}$$
$$(510-490) - 2.57.\ 2.74 < (\mu_1-\mu_2) < (510-490) + 2.57.2.74$$
$$20 - 7.04, < (\mu_1-\mu_2) < 20+7.04$$
$$12.96 < (\mu_1-\mu_2) < 27.04$$

وبناء على ذلك فإنه وبدرجة ثقة ٩٩٪ تكون أجزاء المحراث المصنوعة بطريقه البثق أقوى في المتوسط بمقدار يتراوح بين ١٢.٩٦ و ٢٧.٠٤ مقارنة بطريقة استخدام آلة القوالب والتلحيم.

العينات الصغيرة المستقلة Small, Independent Samples

عندما تكون العينات موضع الدراسة صغيره ويتم سحبها بطريقه مستقلة عن بعضها فإن التوزيع المناسب في هذه الحالة هو توزيع t والذي يستخدم لدراسة المتغير العشوائي الذي هو عبارة عن الفرق بين متوسطات المجتمع.

وفي هذه الحالة فإن العينات الصغيرة لا تعطي تقدير ثابت للتباين وعليه فإننا

نستخدم في هذه الحالة جميع البيانات المتحصل عليها لتقدير التباين حيث يتم استخدام الصيغة الرياضية الموضحة في المعادلة رقم (7.6) بحيث يتم ترجيح التباين باستخدام درجات الحرية[1] :

$$S^2 = \frac{(n_1-1)S_1^{\,2} + (n_2-1)S_2^{\,2}}{n_1+n_2-2} \tag{7.6}$$

وللحصول على الانحراف المعياري لفرق المتوسطات $S_{\bar{X}_1-\bar{X}_2}$ يتم قسمة التباين المتحصل عليه من المعادلة رقم (7.6) باستخدام بيانات كل عينه ثم جمع الناتج وإيجاد الجذر التربيعي كما في المعادلة رقم (7.7) .

$$S_{\bar{x}_1-\bar{x}_2} = \sqrt{S^2/n_1 + S^2/n_2}. \tag{7.7}$$

وللحصول على فتره الثقة بدرجه حرية $(1-\alpha)100$ باستخدام توزيع t يمكن عرض ذلك كالتالي :

$$\begin{gathered} P(-t_{(\alpha/2,\,v)} < t < t_{(\alpha/2,\,v)}) = (1-\alpha) \\ t = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S^2}{n_1}+\dfrac{S^2}{n_2}}} \end{gathered} \tag{7.8}$$

وبالتعويض عن قيمة t في معادلة فتره الثقة وإعادة ترتيب المعادلة فإن فتره الثقة لفرق متوسطات المجتمع $(\mu_1 - \mu_2)$ والتي نرغب في إيجادها تكون كما في المعادلة رقم (7.9) التالية :

$$(\bar{X}_1-\bar{X}_2) - t_{(\alpha/2,\,v)}\cdot S_{\bar{x}_1-\bar{x}_2} < (\mu_1-\mu_2) < (\bar{X}_1-\bar{X}_2) + t_{(\alpha/2,\,v)}\cdot S_{\bar{x}_1-\bar{x}_2} \tag{7.9}$$

ويمكننا إيضاح كيفية تطبيق ذلك من خلال المثال التالي :

قام أحد مربي الحيوانات بدراسة لمعرفة تأثير إضافة فيتامين لمياه الشرب على الوزن المكتسب للحيوان فاختار عشرة حيوانات ووضع الفيتامين في مياه الشرب وأختار مجموعه أخرى حجمها ٩ حيوانات كمجموعة تحكم دون أن يضيف لها فيتامين. وقد لاحظ المربي بأن متوسط الوزن المكتسب من الولادة إلى الفطام للحيوانات المضاف لمياه الشرب الخاصة بها فيتامين يساوي ٢٥ رطلاً بانحراف معياري ٨ رطل بينما متوسط الوزن المكتسب لمجموعة التحكم التي لم يضاف الفيتامين للماء يساوي ٢٠ رطلاً بانحراف معياري ١١ رطلاً.

المطلوب إيجاد فتره الثقة لمتوسط الفرق في الوزن المكتسب للمجموعتين بدرجه ثقة ٩٥٪.

لإيجاد فتره الثقة نحسب أولاً درجه الحرية كالتالي :

$$\nu = (n_1 + n_2 - 2) = (10 + 9 - 2) = 17$$

وحيث إن مستوى الثقة المطلوب ٩٥٪ فإن $\alpha = 0.05$ ومن الجدول رقم (٨) بالملحق نجد أن قيم t هي ± ٢,١١ وعليه يمكن حساب قيمة التباين التجميعي كما يلي.

$$S^2 = [(n_1 - 1)S_1^{\ 2} + (n_2 - 1)S_2^{\ 2}] \ / \ (n_1 + n_2 - 2)$$
$$S^2 = [(10-1)8^2 + (9-1)11^2]/(10+9-2)$$
$$S^2 = [576 + 968] / 17 = 90.82$$

وباستخدام القيمة المتحصل عليها أعلاه يمكن حساب الخطأ المعياري لفرق المتوسطات.

$$S_{\bar{x}_1-\bar{x}_2} = \sqrt{S^2/n_1 + S^2/n_2}$$
$$S_{\bar{x}_1-\bar{x}_2} = \sqrt{90.82/10 + 90.82/9}$$
$$S_{\bar{x}_1-\bar{x}_2} = \sqrt{9.082 + 10.091} = \sqrt{19.173}$$
$$S_{\bar{x}_1-\bar{x}_2} = 4.38$$

وأخيراً يمكن التعبير عن فتره الثقة للفرق بين المتوسطات باستخدام المعادلة رقم (7.9) والتعويض بالقيم المتحصل عليها أعلاه في ذلك كالتالي :

$$(\bar{X}_1 - \bar{X}_2) - t_{(\alpha/2,v)} \cdot S_{\bar{x}_1-\bar{x}_2} < (\mu_1 - \mu_2) < (\bar{X}_1 - \bar{X}_2) + t_{(\alpha/2,v)} \cdot S_{\bar{x}_1-\bar{x}_2}$$
$$(25 - 20) - 2.11 \cdot 4.38 < (\mu_1 - \mu_2) < (25 - 20) + 2.11 \cdot 4.38$$
$$5 - 9.24 < (\mu_1 - \mu_2) < 5 + 9.24$$
$$-4.24 < (\mu_1 - \mu_2) < 14.24$$

ونلاحظ أن قيمة فترة الثقة تشمل قيمة الصفر حيث إنها تمتد من -٤,٢٤ إلى ١٤,٢٤ وهذا يعني أنه قد لا يكون هنالك فرق في الوزن المكتسب للحيوانات التي تم تغذيتها بالفيتامين مقارنة بحيوانات التحكم ؛ نظراً لأن $0 = (\mu_1 - \mu_2)$ مشمولة بفترة الثقة.

وبناء على هذه النتائج للتجربة ، فإن منتج الحيوانات قد لا يكون لديه الرغبة في شراء الفيتامين لحيواناته. وهذه الفترة كبيرة نسبياً ويحدث ذلك في حالة أن حجم العينة صغير والتباين كبير. ويمكن إجراء بعض الخطوات لتقليل التباين وإعادة التجربة إذا كان ذلك ممكناً.

العينات الصغيرة المزدوجة (غير المستقلة) Small, Paired Samples

إذا كان لدينا عينات صغيرة الحجم وبيانات مترابطة ، مثلاً البيانات غير مستقلة ؛ نظراً لوجود بعض العوامل التي تجعلها مترابطة ، في هذه الحالة نستخدم

توزيع t كقاعدة لحساب فترة الثقة بدرجه ثقة $100\,(1-\alpha)$ لفرق متوسطات المجتمع $(\mu_1 - \mu_2)$ ، وهذا يعني أننا ننشئ فتره الثقة باستخدام البيانات التي تمثل الفرق بين المشاهدات للعينتين بدلاً من استخدام البيانات الأصلية نفسها. ويحدث اقتران البيانات بعضها ؛ نظراً لنوع التجربة التي تم إجراؤها. في هذه الحالة ، فإن التجربة تشتمل على مشاهدات في نفس المجال قبل وبعد إجراء التجربة ، أو قد تتأثر المشاهدات في العينتين ببعض العوامل التي لا يمكن التحكم فيها في حالة التصميم العشوائي ، مثل نوع التربة في حالة تجربة إنتاجية القطن. وفي هذه الحالة فإن حجم العينات يجب أن يكون متساوياً حتى نستطيع ربطهما وعليه إيجاد الفرق ، ويمكن التغلب على العامل المسبب لارتباط العينات ببعضها عن طريق طرح قيم العينات من بعضها فمثلاً الوراثة تؤثر على قدرة الحيوانات على اكتساب الوزن. وعليه فإنه عند إجراء الدراسات المتعلقة بالوزن المكتسب يفضل استخدام الحيوانات التي بينها ترابط مثل التوائم وإجراء اختبار t على النتائج.

ويمكن اشتقاق فتره الثقة باستخدام الصيغة الاحتمالية التالية :

$$P(-t_{(\alpha/2,v)} < t < t_{(\alpha/2,v)} = (1-\alpha)$$

ولكن في هذه الحالة فإن الصيغة الرياضية المستخدمة لحساب قيمة t مختلفة. ويمكن توضيح ذلك بتعريف المتغير العشوائي d والذي يمثل الفرق بين بيانات العينة ، أي $d_1 = x_{11} - x_{21}$ ، $d_2 = x_{12} - x_{22}$ .... ، $d_n = x_{1n} - x_{2n}$ لجميع بيانات العينة ، و $\bar{d}$ يعبّر عن متوسط الفروق ، و S_d^2 يعبّر عن تباين الفروق. ويمكن حساب هذه المتغيرات مثل أي حساب لعينه مفرده فمثلاً $\bar{d} = \sum d/n$ ، إلى آخره. ويمكن كتابة الصيغة الرياضية لـ t كالتالي :

$$t = \frac{\bar{d} - (\mu_1 - \mu_2)}{\frac{S_d}{\sqrt{n}}} \qquad (7.10)$$

وبالتعويض بالطرف الأيمن من معادلة t السابقة في صيغة فترة الثقة السابقة وإجراء بعض التعديلات نحصل على فترة الثقة لـ $(\mu_1 - \mu_2)$ والموضحة في المعادلة رقم (7.11) كالتالي :

$$\bar{d} - t_{(\alpha/2,\, v)} \cdot S_d / \sqrt{n} < (\mu_1 - \mu_2) < \bar{d} + t_{(\alpha/2,\, v)} \cdot S_d / \sqrt{n} \qquad (7.11)$$

ويمكن إيضاح كيفية تطبيق ذلك من خلال المثال التالي :

لاختبار هجينين جديدين من حبوب الذرة المزروعة تحت ظروف زراعة أولية ، قامت شركة البذور باختيار عدد ٩ مزارع عشوائياً وتم إعطاؤها للمزارع لزراعتها في أحواض التجربة. وتم قياس الإنتاجية بالقنطار / أيكر للمواقع التسعة والموضحة في الجدول رقم (٧,٢).

الجدول رقم (٧,٢). إنتاجية حبوب الذرة بالقنطار/أيكر لعدد تسع مزارع مختلفة.

المزرعة	هجين ١	هجين ٢
١	٨٥	٧٩
٢	٨٨	٨٠
٣	٥٨	٦٠
٤	٩٤	٩٢
٥	٨٥	٧٨
٦	٩٣	٨٧
٧	٧٤	٧٥
٨	٨٠	٧٦
٩	١٠١	٩٥

المطلوب إيجاد فترة الثقة لمتوسط فرق الإنتاجية للهجينين بدرجة ثقة ٩٥٪.

للحصول على فتره الثقة بدرجة ٩٥٪ نحسب أولاً درجة الحرية كالتالي :

$$v = (n-1) = (9-1) = 8$$

ثم نوجد قيم t عند مستوى معنوية $\alpha = 0.05$ من الجدول رقم (٨) بالملحق والتي تساوي ±٢,٣٠٦ . ثم نحسب قيمة $\bar{d}$ و S_d بحساب d أولاً وإجراء الحسابات اللازمة باستخدام الصيغة الرياضية الخاصة بها والموضحة في الجدول رقم (٧,٣) وهي :

$$\bar{d} = \Sigma d/n = 36/9 = 4$$

$$S_d^{\,2} = \Sigma(d - \bar{d})^2/(n-1) = 102/(9-1) = 12.75$$

$$S_d = \sqrt{S_d^{\,2}} = \sqrt{12.75} = 3.57$$

الجدول رقم (٣.٧). حسابات متوسط وتباين الفرق d لبيانات إنتاجية الذرة الهجين.

المزرعة	هجين ١	هجين ٢	d	$d-\bar{d}$	$(d-\bar{d})^2$
١	٨٥	٧٩	٦	٦- ٤=٢	٤
٢	٨٨	٨٠	٨	٨- ٤=٢	١٦
٣	٥٨	٦٠	- ٢	- ٢- ٤=- ٦	٣٦
٤	٩٤	٩٢	٢	٢- ٤=- ٢	٤
٥	٨٥	٧٨	٧	٧- ٤=٣	٩
٦	٩٣	٨٧	٦	٦- ٤=٢	٤
٧	٧٤	٧٥	- ١	- ١- ٤=- ٥	٢٥
٨	٨٠	٧٦	٤	٤- ٤=٠	٠
٩	١٠١	٩٥	٦	٦- ٤=٢	٤
الإجمالي			٣٦	٠	١٠٢

وبالتعويض عن القيم نحصل على فترة الثقة، كما في الخطوات التالية :

$$\bar{d} - t_{(\alpha/2, v)} + S_d/\sqrt{n} < (\mu_1 - \mu_2) < \bar{d} + t_{(\alpha/2, v)} \cdot S_d / \sqrt{n}$$
$$4 - 2.306 \cdot 3.57/\sqrt{9} < (\mu_1 - \mu_2) < 4 + 2.306 \cdot 3.57/\sqrt{9}$$
$$4 - 2.74 < (\mu_1 - \mu_2) < 4\ 2.74$$
$$1.26 < (\mu_1 - \mu_2) < 6.74$$

وتشير فترة الثقة المتحصل عليها أنه بدرجة ثقة ٩٥٪ فإن متوسط الإنتاجية للهجين الأول أكبر من الهجين الثاني بمقدار يتراوح بين ١,٢٦ و ٦,٧٤ .وباستخدام هذه الطريقة تم إلغاء أثر الفروقات الفردية من مزرعة إلى مزرعة والتي تؤثر على مستوى الإنتاجية مثل سقوط الأمطار، اختلاف خصوبة التربة، وكذلك المقدرة الإدارية للملاّك.

فترات الثقة للفرق بين نسبتين $p_1 - p_2$

Confidence intervals for $p_1 - p_2$

يمكن إيجاد فترة الثقة بدرجة ثقة $100\,(1-\alpha)$ للفرق بين النسب للعينات باستخدام تقريب التوزيع الطبيعي لتوزيع المعاينة ذي الحدين للمتغير العشوائي $(\hat{P}_1 - \hat{P}_2)$ كمقدّر للفرق بين نسب المجتمع $(P_1 - P_2)$ ، حيث $\hat{P}_1$ تعرف على أنها r_1/n_1 و $\hat{P}_2$ تعرف على أنها تساوي r_2/n_2. وتكون صيغة فترة الثقة ممثلة بالتالي :

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = (1 - \alpha)$$

و Z معرّفه بالمعادلة رقم (7.12) التالية :

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{S_{p_1 - p_2}} \qquad (7.12)$$

يمكن الحصول على الثقة $(P_1 - P_2)$ بالتعويض عن الطرف الأيمن من معادلة Z آنفاً في صياغة فترة الثقة وإعادة ترتيب الحدود لنحصل على المعادلة رقم (7.13) والتي تعبّر عن فترة الثقة.

$$(\hat{p}_1 - \hat{p}_2) - Z_{\alpha/2} \cdot S_{p_1-p_2} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + Z_{\alpha/2} \cdot S_{p_1-p_2} \qquad (7.13)$$

ويمكن إيضاح تطبيق الصيغة الرياضية أعلاه من خلال المثال التالي :

ترغب إحدى المختصات بالإنتاج النباتي في تقدير الفرق في نسبة نمو عقل العنب في البيوت المحمية. قامت بمعاملة مجموعة من النباتات عددها ١٦٠ بمنشط للجذور ولاحظت أن ١٤٠ نباتاً تبقى حية في نهاية التجربة، بينما تركت مجموعة أخرى من النباتات عددها ١٢٠ بدون معاملة ولاحظت أن ٩٦ منها تبقى حية. المطلوب إيجاد فترة الثقة بدرجة ثقة ٩٥٪ لنسبة الفرق في النباتات الحية لهذه ا لتجربة.

لإيجاد فترة الثقة فإننا أولاً يجب أن نحسب قيمة Z من الجدول رقم (٧) بالملحق عند قيمه $\alpha = 0.05$ حيث نجد أن $Z_{\alpha/2} = \pm 1.96$. ثم نوجد الحسابات التالية :

$$\begin{aligned}
\hat{p}_1 &= r_1/n_1 = 140/160 = 0.875; \\
\hat{p}_2 &= r_2/n_2 = 96/120 = 0.800 \\
S_{p_1-p_2} &= \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \\
S_{p_1-p_2} &= \sqrt{(0.875)(1-0.875)/160 + (0.800)(1-0.800)/120} \\
S_{p_1-p_2} &= \sqrt{0.1094/160 + 0.1600/120} \\
S_{p_1-p_2} &= \sqrt{0.00068 + 0.00133} \\
S_{p_1-p_2} &= \sqrt{0.002} = 0.04
\end{aligned}$$

وعليه فإن لدينا الآن معلومات كافيه لإيجاد فترة الثقة باستخدام المعادلة رقم (7.13) كالتالي :

$$(\hat{p}_1 - \hat{p}_2) - Z_{\alpha/2} \cdot S_{p_1 - p_2} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + Z_{\alpha/2} \cdot S_{p_1 - p_2}$$
$$(0.875 - 0.800) - 1.96\,(0.04) < (p_1 - p_2) < (0.875 - 0.800) + 1.96 \cdot (0.04)$$
$$0.075 - 0.078 < (p_1 - p_2) < 0.075 + 0.078$$
$$-0.003 < (P_1 - P_2) < 0.153$$

وحيث إن فترة الثقة تحتوي على الصفر فإن المختصة ربما تحتاج لمعلومات إضافية قبل استخدام المنشط الجذري. وتشير النتائج آنفاً إلى أن الفرق في نسبة نمو عقل العنب المعاملة بالمنشط الجذري يتراوح بين −٠,٠٠٣ و ٠,١٥٣ مقارنة بالنباتات التي لم تعامل بالمنشط الجذري.

خاتمة Endnote

(١) يمكن عرض صيغة إجمالي درجات الحرية كالتالي:

$(n_1 - 1) + (n_2 - 1)$ يمكن كتابتها على النحو $(n_1 + n_2 - 2)$

تمارين Exercises

١- في عينة عشوائية بسيطة لعدد ١٠٠ حيوان من سلالات مختلطة وجد أن متوسط وزن البيع يساوي ٢٢٠ رطلاً بانحراف معياري يساوي ٢٠ رطلاً بينما في عينه لعدد ٤٠٠ حيوان من سلالات نقيه وجد أن متوسط وزن البيع يساوي ٢١٠ رطل بانحراف معياري يساوي ٣٠ رطل. المطلوب إيجاد فترة الثقة بدرجه ثقة ٩٥٪ للفرق الحقيقي في متوسط وزن البيع للحيوانات في المجموعتين.

٢- يرغب أحد المزارعين في شراء أسلاك بالات لمزرعته فقام بأخذ عينة من أسلاك ربط متوسطة القوه حجمها ٥٠ قطعة وعينة أخرى من أسلاك ربط قويه حجمها ٦٠ قطعة وأختبرها في أحد المعامل القريبة لمعرفة قوة الشد للأسلاك

فكان تقرير المعمل كالتالي :

أ) متوسط قوة الشد للأسلاك متوسطة القوه يساوي ٩٥,٨ رطلاً بانحراف معياري يساوي ٥ رطل.

ب) ٣٤٪ من الأسلاك متوسطة القوه لها قوة شد تساوي ١٠٠ رطل أو أكثر.

ج) متوسط قوة الشد للأسلاك القوية تساوي ٩٩ رطلاً بانحراف معياري يساوي ٦ رطل.

د) ٣٠٪ من الأسلاك القوية لها قوة شد تساوي ١٠٠ رطل أو أكثر والمطلوب :

- أوجد فتره الثقة بدرجه ثقة ٩٠٪ لمتوسط الفرق في قوة الشد لكلا النوعين من الأسلاك.
- أوجد فترة الثقة بدرجه ثقة ٩٥٪ للفرق الفعلي في نسبة السلك والتي لها قوة شد ١٠٠ أو أكثر لكلا النوعين.
- فسّر النتائج المتحصل عليها.

٣- في دراسة لتقييم معدل الإنبات لبذور غار الجبل تم اختيار ١٠٠ بذره وكشطها ثم زراعتها واختيار ٧٠ بذرة أخرى ومعاملتها بالحامض ثم زراعتها. وكانت النتائج أن ٥٠ بذرة من البذور المكشوطة نبتت بينما نبت ٤٨ بذرة من البذور المعاملة بالحامض. أوجد فتره الثقة للفرق في نسبة الإنبات لنوعي المعاملة التي تمت على البذور.

٤- في عينة عشوائية لعدد ١٠ مزارع بطيخ وجد أن متوسط أجر الساعة لعمال الحصاد يساوي ٧,٢٥ دولار بانحراف معياري ١,٢٥ دولار ، بينما متوسط أجر الساعة للعمال في ١٢ مزرعة فواكه يساوي ٨ دولارات بانحراف معياري يساوي ١

دولار. أوجد فتره الثقة بدرجه ثقة ٩٠٪ لمتوسط الفرق في أجر الساعة المدفوع لعمال الحصاد لكلا نوعي المزارع.

٥- في عينة عشوائية لعدد ٨ من مزارعي القمح وجد أن متوسط تكاليف البيع تساوي ٠,١دولار/ بوشل بانحراف معياري يساوي ٠,٠٥ دولار/بوشل ، بينما وجد في عينة عشوائية لعدد ١٥ من مزارعي الشعير أن متوسط تكاليف البيع تساوي ٠,٠٧ دولار/بوشل بانحراف معياري يساوي ٠,٠٣ دولار/بوشل. أوجد فتره الثقة بدرجه ثقة ٩٥٪ لمتوسط الفرق في تكاليف البيع للبوشل.

٦- تمت تغذية عدد ٦ مجموعات من توأم أبقار الحليب مع صغارهن بنفس الحصص ماعدا أحد التوأم أعطي هرمون BST والذي يزيد من إنتاج الحليب. المطلوب إيجاد فتره الثقة بدرجه ثقة ٩٥٪ للفرق في إنتاج الحليب بالرطل للحيوان للبيانات التالية :

التوأم	كمية الحليب بالرطل بدون BST	كمية الحليب بالرطل بـــ BST
١	٤٥	٤٨
٢	٥٠	٥١
٣	٦٠	٦٤
٤	٥٥	٥٦
٥	٥٣	٥٧
٦	٤٩	٥٢

٧- البيانات التالية توضح الإنتاجية المشتركة لصنفين من الشعير. كل صنف تمت زراعته في مزرعة مختلفة. المطلوب تقدير الفرق في الإنتاجية بدرجه ثقة ٩٥٪ لفترة الثقة.

الصنف	إنتاجيه عاليه	صنف أوكالا ٩٥
١	٣٧	٤٢
٢	٤١	٤٥
٣	٣٥	٣٣
٤	٥٢	٥٦
٥	٤٤	٥٠
٦	٣٠	٣٢
٧	٣٨	٣٥
٨	٥٤	٥٥

الفصل الثامن

اختبار الفروض
Hypothesis Testing

توضح الخطوات التي تم عرضها في الفصل السابق التقدير بنقطة والتقدير بفترة لمعالم المجتمع ، وذلك باستخدام فترة ثقة للتعبير عن درجه عدم التأكد حول القيمة الحقيقية للمعالم. وكذلك تساهم الفترات في اختبار القيم الحقيقية المفترضة لمعالم المجتمع. وفي الحقيقة فإننا مهتمون بهذه القيم والتي تسمى الفروض الإحصائية. تعرّف الفروض الإحصائية بأنها عبارة عن افتراض يتم عمله حول بعض معالم أو متغيرات المجتمع.

وفي الحقيقة فإنه من الصعوبة تصنيف كل الفروض إحصائياً. فمثلاً ، العبارة التي تقول إن " ماشية البرانفيه Braunvieh هي أفضل لحوم البقر في الولايات المتحدة " ليست عبارة إحصائية.ولكن بتعديل بسيط يمكن تحويلها إلى فرض إحصائي. حيث إن العبارة " البرانفيه Braunvieh هي أفضل قطيع أبقار مفضل من قبل منتجي اللحوم بالولايات المتحدة " عبارة إحصائية ؛ نظراً لأنها تعبّر عن فرض حول المعلمة التي تعبر عن نسبة مربي لحوم البقر في الولايات المتحدة والتي يمكن اختبارها باستخدام البيانات.

أنواع الاختبارات Types of Tests

لتحديد ما إذا كان الفرض الإحصائي صحيح بتأكد تام فإن ذلك ربما يتطلب منا اختبار كامل المجتمع. ونظراً لأن ذلك عموماً غير عملي لعدد من الأسباب، فإننا تقريباً نأخذ عينة ونجري اختبار الفرض الإحصائي باستخدام بيانات العينة. ونحن فعلاً نعلم من النقاش السابق بأن المعاينة تحتوي على أخطاء محتملة. لذا، فإنه عند اختبار الفرض الإحصائي فإنه يتم تحديد قيمه للخطأ التي يمكن السماح بها وعادة يتم تحديد تلك القيمة بحيث تكون أصغر ما يمكن. إضافة لما سبق، تختلف الفروض الإحصائية في درجة تعقيدها اعتماداً على نوعية الفرض هل هو بسيط أم مركب. فالفرض البسيط يكون عبارة عن اختبار قيمة واحدة للمعلمة، مثلاً $H_0 : \mu = 75$ فرض بسيط حول متوسط المجتمع. وفي العادة يتم استخدام الرمز H_0 للتعبير عن الفرض المراد اختباره. وتشير العبارة بعد النقطتين إلى الفرض الفعلي. وعلى العكس، فإن الفرض المركب يحتوي مدى من القيم للمعلمة المراد اختبارها. وذلك المدى ربما يتكون من رقم صغير أو تقريباً مجموعة لا نهائية من القيم. ولاختبار الفروض المركبة فإنه يجب تحديد ما إذا كانت المعلمة تأخذ أي من القيم المحتملة في المدى أم لا. ومن الأمثلة على الفروض المركبة $H_0 : u < 150$ و $H_0 : \mu_1 - \mu_2 > 0$.

وعند صياغة الفرض الإحصائي فإننا نحدّد عبارتين ونقرر أي منهما يمكن اختياره. وهذه العبارتان أحداهما تدعى فرض العدم والأخرى تسمى الفرض البديل يرمز لفرض العدم بالرمز الرياضي H_0 ويشير هذا الفرض لعدم وجود اختلاف، أو عدم وجود أثر، أو لا شيء. بينما الفرض البديل يرمز له بالرمز $\boldsymbol{H_a}$ أو H_1 (حيث تستخدم في العادة كلا الرمزين) وهي تعبر عن العبارة الإحصائية لقيمة معلمه المجتمع التي نعتقد أنها صحيحة. ولذلك، عند اختبار الفروض فإننا نرغب في رفض

فرض العدم (ندعي عادة أنه فرض خاطئ ونقبل الفرض البديل كفرض صحيح)، وفي الحالات التي لا يمكن رفض فرض العدم، فإننا لا نقول إنها صحيحة ولكن نقول لا يمكن رفضها بسبب أنه يمكن رفضها عند إجراء بعض الاختبارات المستقبلية. وفي التحقيق العلمي، فإنه دائماً هناك احتمالية أن تسبب البيانات الجديدة رفض الفرض كفرض خاطئ والتي كنّا نعتقد بأنه صحيح.

فرض العدم والفرض البديل قد تكون بسيطة أو مركبه أو مزيج من الاثنين. فمثلاً يمكن اختبار فرض العدم البسيط $H_0 : P = 0.166$ مقابل الفرض البديل البسيط $H_a : P = 0.25$ أو مقابل الفرض البديل المركب $H_a : P \neq 0.166$. من جهة أخرى يمكن اختبار فرض العدم المركب $H_0 : \mu < 10$ مقابل فرض بديل بسيط مثل $H_1 : \mu = 10$ أو مقابل فرض بديل مركب $H_1 : \mu \geq 10$. وأياً كانت صياغة الفرض، فإن القيمة الصحيحة للمعلمة يجب أن تكون في المجموعة المحددة لـ H_0 أو المجموعة المعرّفة بـ H_a. أي أن فرض العدم والفرض البديل يجب أن تحتوي على كل القيم المحتملة التي يمكن أن تأخذها المعلمة. وبالتالي يمكن عمل ذلك بجعل المجموعتين متكاملتين فمثلاً إذا كان فرض العدم $H_0 : \mu > 50$ فإن الفرض المكمل له هو $H_a : \mu \leq 50$. فالفرض في المثال السابق يعني إجراء اختبار من طرف واحد؛ لأنه إذا كان فرض العدم غير صحيح فإن جميع القيم للفرض البديل تقع في طرف واحد منه (الطرف الأيسر في هذه الحالة)، ولذا فإن هناك اختبار طرف واحد للطرف الأيسر واختبار طرف واحد للطرف الأيمن اعتماداً على القيم المحدّدة في الفرض البديل هل هي أقل من أو أكبر من تلك القيم في فرض العدم.

أما في حالة استخدام فرض عدم بسيط وفرض بديل مركب فإن الاختلاف في قيم المعلمة يمكن أن يكون في أحد الطرفين من قيم فرض العدم وبذلك نستخدم اختبار الطرفين.

فمثلاً إذا كانت لدينا العبارة $H_0 : P = 0.125$ مقرونة بالفرض البديل $H_a : P \neq 0.125$ فهذا يعني استخدام اختبار من طرفين.

وتتمثل المشكلة في اختبار الفروض في كيفية استخدام بيانات العينة لعمل اختبار بين عبارتين متنافيتين حول قيمة المعلمة. وكما تمت الإشارة له سابقاً فإن مثل هذا القرار قد يشتمل على خطأ. وعلى وجه التحديد، فإن هناك نوعين من الأخطاء المحتملة. القرار بقبول الفرض البديل عندما يكون فرض العدم صحيح (خطأ النوع الأول)، أو القرار بقبول فرض العدم عندما يكون الفرض البديل صحيح (الخطأ من النوع الثاني).

أخطاء النوع الأول والنوع الثاني Type I and Type II Errors

لعرض أخطاء النوع الأول والثاني نفترض على سبيل المثال أن أحد المختصين في الاقتصاد الزراعي يرغب في اختبار متوسط إنفاق الأسرة السنوي على الفواكه والخضار الطازجة. قام الاقتصادي بصياغة الفرض الإحصائي بحيث إن فرض العدم يتمثل في أن متوسط الإنفاق أقل من أو يساوي ١٢٠٠دولار $H_0 : \mu \leq 1200$ مقابل الفرض البديل أن متوسط الإنفاق أكبر من ١٢٠٠دولار $H_a : \mu > 1200$ فيكون الباحث قد ارتكب خطأ من النوع الأول إذا أدت النتائج إلى قبول الفرض البديل عندما يكون الإنفاق الفعلي يساوي ١٢٠٠دولار أو أقل، أو خطأ من النوع الثاني إذا لم يرفض فرض العدم H_0 عندما يكون الإنفاق أكبر من ١٢٠٠دولار.

الجدول رقم (٨.١). أخطاء النوع الأول والنوع الثاني.

القرار	الحالة	
	H_0 صحيح	H_a صحيح
قبول H_0	قرار سليم	خطأ من النوع الثاني باحتمال β
رفض H_0	خطأ من النوع الأول باحتمال α	قرار سليم

ونظراً لاحتماليه الوقوع في خطأ من النوع الأول أو النوع الثاني في معظم الاختبارات الإحصائية، فإننا نسعى لتصميم نموذج لأخذ تلك الأخطاء في الحسبان. ولعمل ذلك يجب الاهتمام باحتمالية وقوع تلك الأخطاء. في الفصل السابق كان الحرف الإغريقي α يشير إلى احتمالية عدم شمول فترة الثقة للقيمة الحقيقة لمعلمة المجتمع وقد تم اختيار تلك القيمة بحيث تكون صغيرة. أما في هذا الفصل فإننا نستخدم α لعرض أقل احتمال للوقوع في الخطأ الأول ويتم اختيارها أيضاً. من جهة أخرى فإن احتمال الوقوع في الخطأ من النوع الثاني يعبّر عنه بالحرف الإغريقي بيتا (β). ويوضح الجدول رقم (٨،١) عرض للحالات التي تؤدي للوقوع في تلك الأخطاء. ولعرض طريقه حساب α و β نفترض أننا نرغب في اختبار كفاءة حجر النرد. نفترض وجود حجر نرد غير سليم (مرجح) وقد علمنا أن احتمال ظهور الرقم واحد يساوي $P = 0.25$ وقد تم خلط حجر النرد المرجح من غير قصد مع مجموعة من أحجار النرد الأخرى السليمة. وحيث إنه لا يوجد علامة تحدد أو تعرّف ذلك الحجر، سيتم اختيار الحجر عشوائياً ورميه ١٠٠ مره $n = 100$ وملاحظة ظهور العدد رقم واحد والذي يمكن أن نرمز له بالرمز r ويمكن حساب نسبة ظهور العدد واحد من خلال المعادلة رقم (8.1).

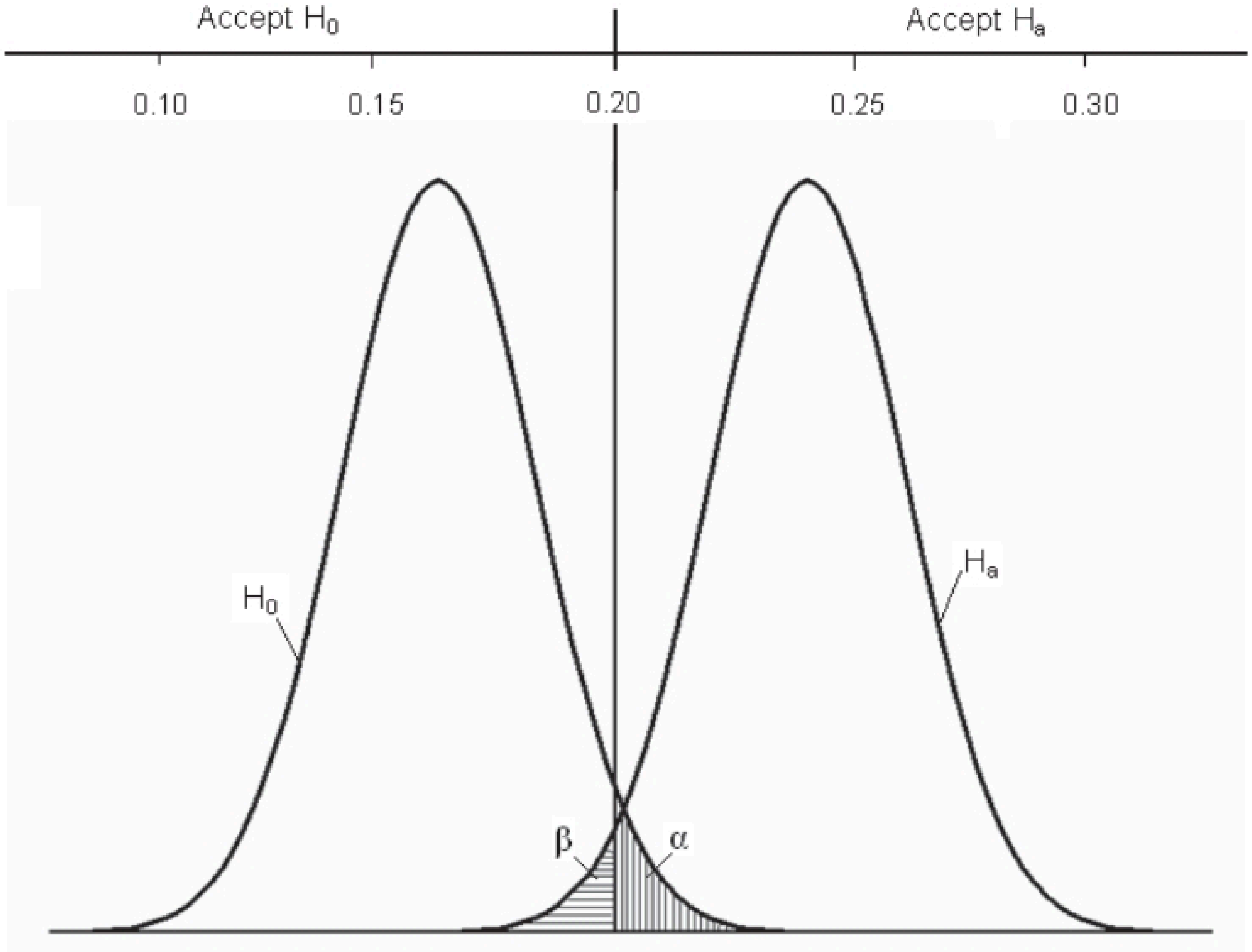
$$\hat{P} = r/n \tag{8.1}$$

فإذا كانت قيمة $\hat{P}$ ليست أكبر من قيمه مناسبة تم تحديدها فإننا سوف نقبل نتيجة الاختبار بأن الحجر كفء أما إذا كانت قيمه $\hat{P}$ أكبر من قيمتنا ، فإننا سنرفض الفرض القائل بكفاءة الحجر وسوف نحصل على حجر النرد المرجح. ولذلك فإن لدينا قاعدة لاتخاذ القرار للتحكم في الخطوات. والقيمة المناسبة التي تم اختيارها تدعى القيمة الحرجة. لنفترض أننا حددنا القيمة الحرجة بـ $P = 0.20$ مثلاً ، فإذا لم نحصل على الرقم ١ عشرين مرة من ١٠٠ رمية للحجر ، فإننا لن نرفض فرض العدم ونقبل بكفاءة الحجر ولكن إذا حصلنا على الرقم ١ أكثر من عشرين فإننا سوف نرفض فرض العدم مقابل الفرض البديل ونقول بعدم كفاءة الحجر. ولذلك فإن لدينا الفرض.

$$H_0 : P = 0.167$$
$$H_a : P = 0.250$$

وقاعدة اتخاذ القرار هي $P = 0.20$. وسيتم عرض ذلك باستخدام تقريب التوزيع الطبيعي لتوزيع ذي الحدين حيث إن المتوسط هو $\mu = p$ والانحراف المعياري يمكن حسابه من $\sigma = \sqrt{Pq/n}$. وفي الحقيقة فإن لدينا توزيعين أحدهما عندما يكون فرض العدم صحيح $(\mu = 0.167 , \sigma = 0.037)$ والآخر عندما يكون الفرض البديل صحيح $(\mu = 0.25 , \sigma = 0.043)$.

ويوضح الشكل رقم (٨،١) توزيعات المعاينة لفرض العدم والفرض البديل لتجربة حجر النرد.



الشكل رقم (٨.١). توزيعات المعاينة لفرض العدم والفرض البديل لحجر النرد.

حيث توضح المساحة المظللة الخطوط الرأسية وتحت التوزيع الأول قيمة α والتي تقع يمين قاعدة القرار $P = 0.20$ ، بينما توضح المساحة المظللة بالخطوط الأفقية تحت التوزيع الثاني قيمة β والتي تقع يسار قاعدة القرار $P = 0.20$. ويجب ملاحظة أنه لابد من تحديد قاعدة اتخاذ القرار قبل إجراء الاختبار وكذلك قبل معرفة قيم α ، β . والآن يمكننا استخدام الصيغة الرياضية لـ Z لحساب الاحتمالات الفعلية للخطأ الأول والثاني. ويمكن إعادة كتابة صيغة حساب الاحتمالات باستخدام z كالتالي :

$$Z = (\hat{P} - P) / \sqrt{Pq/n} \qquad (8.2)$$

وفي مثالنا الحالي فإن $\hat{P} = 0.2$ كقيمة لاتخاذ القرار . وبحساب قيمة Z_α

نستخدم القيم المصاحبة لتوزيع المعاينة عندما تكون H_0 صحيحة، أي $P = 0.167$ و $\sigma = 0.037$ لذا فإن Z_α تساوي :

$$Z_\alpha = (0.20 - 0.167) / 0.037 = 0.89$$

وباستخدام توزيع Z من الجدول رقم (٧) بالملحق نجد أن احتمال أن تزيد قيمة Z عن ٠,٨٩ هي :

$$0 \cdot 500 - 0.3133 = 0.1687 = \alpha$$

ويمكن لهذا الاختبار أيضاً حساب احتمال الوقوع في الخطأ من النوع الثاني باستخدام صيغه Z بالقيم المصاحبة للتوزيع عندما تكون H_a صحيحة حيث :

$$Z_\beta = (0.20 - 0.25) / 0.043 = -1.16$$

وباستخدام توزيع Z من الجدول رقم (٧) بالملحق نجد أن احتمال أن تقل قيمة Z عن 1.16- هي

$$0.5000 - 0.3770 = 0.1230 = \beta$$

وحيث إن القيمة الحرجة في قاعدة القرار هي التي تحدد في النهاية حجم الخطأ الأول والخطأ الثاني عند صياغة فرض العدم والفرض البديل، لذا فإنه يمكننا تحريك قاعدة اتخاذ القرار لليمين بعيداً عن قيمة H_0 التي تساوي $P = 0.167$ إلى $P = 0.22$ ولذلك نخفض احتمال الخطأ الأول إذا كنا نعتقد أن قيمة α والتي تساوي ٠,١٨٦٧ كبيره جداً بالنسبة لهذا الاختبار.

ولكن باتخاذ هذا الإجراء فإننا قد زدنا احتمال β ؛ نظراً لأن خط اتخاذ القرار

يؤثر على كلا الخطأين ولذلك هناك تبادل بينهما بحيث لا يمكن تقليل احتمال أحد الأخطاء دون زيادة احتمال الخطأ الآخر وستتم مناقشة ذلك لاحقاً.

والآن ما هو تفسير الخطأ الأول في مثال كفاءة حجر النرد السابق حيث إننا رمينا الحجر ١٠٠ مرة وكانت قيمة $\hat{P}$ التي تم حسابها من التجربة أكبر من ٠,٢٠ . ومن ثمَّ بناءً على قاعدة القرار فإننا سنرفض فرض العدم. وبناء على نتيجة هذا الاختبار فإننا حصلنا على أكثر من ٢٠٪ للرقم ١ ولذلك يمكننا القول بأن الحجر غير سليم. وفي الحقيقة فإن الخطأ من النوع الأول يحدث إذا كان الحجر سليماً. وبطريقه ما فإن التجربة أدت إلى قيمة مرتفعة لـ $\hat{P}$ على الرغم من أن الحجر كفء. وهذا يحدث بنسبة $100\,\alpha$ وفقاً للطريقة التي تم بها إجراء الاختبار ، أو ١٨,٦٧٪ في المثال. ولذلك فإننا في هذه الحالة ، قد ارتكبنا خطأ من النوع الأول بالقول إن الحجر السليم معوجاً أو ملتوياً.

وبطريقة مشابهة ، فإن الخطأ من النوع الثاني يمكن أن يحدث. في هذه الحالة ، تم رمي الحجر ١٠٠ مرة وتم حساب قيمة $\hat{P}$ والتي ستكون صغيره ، أقل من ٠,٢ ، وبناء على قاعدة اتخاذ القرار فإننا لم نستطيع رفض فرض العدم ، أي أننا نستنتج أن الحجر سليم. وفي الحقيقة فإن الخطأ من النوع الثاني يقع إذا كان الحجر معطوب. وحيث إن الحجر تم ترجيحه فعلياً لذا فإنه في المتوسط سيكون نسبة ظهور الرقم١ هي $P = 0.25$ ، ولكن في هذا المثال حصلنا على بعض النسب القليلة لظهور الرقم١. وعليه فإن الاختبار يوضّح أن الحجر سليم باحتمال β ، والذي يساوي ٠,١٢٣٠ أو ١٢,٣٪ معظم الوقت. كما أن القيمة $1 - \beta$ والتي تساوي ٠,٨٧٧٠ تسمى قوة الاختبار.

التحكم بقيم α و β

Controlling Both β and α

نظراً لأن كلا الخطأين غير مرغوب فيه وقد تكون احتمالاتها كبيره، فإننا نحاول اختيار قاعدة القرار والتي يمكن أن توازنها. في مثال حجر النرد تم أولاً تحديد القيمة الحرجة لقاعدة اتخاذ القرار ثم حسبنا القيم المصاحبة لكل من α و β. وحيث إن تلك المخاطر مرتبطة بحجم العينة ، وزيادة حجم العينة n يقلل من قيم α و β ، لذا فإن اختيارنا للقيمة الحرجة يعتمد على حجم العينة. وعليه ، فإننا في التطبيق نختار حجم α ، والذي يحدد القيمة الحرجة، وحجم العينة ثم نحسب β لقيم الفرض البديل H_a المهمة. فإذا كانت قيمة β كبيره جداً فإننا نحسب حجم العينة والذي يعطي قيمة مقبولة وزيادة العينة إلى الحجم الناتج إذا كان ذلك ممكناً ويجب ملاحظة أنه يجب أن تتم تلك الحسابات في مرحلة التخطيط لاختبار الفرض.

اختبارات الفروض Tests of Hypotheses

الآن يمكننا مناقشة بعض من الاختبارات المعيارية للفروض لمجتمع أو مجتمعين. أما الاختبارات لثلاثة مجتمعات أو أكثر فستتم مناقشتها لاحقاً. وسنبدأ باختبار الفروض الأوسع انتشاراً وهو اختبار العينة الواحدة التي تحتوي على متوسط المجتمع μ.

ولنجاح الاختبار فإننا يجب أن نتبع خطوات معينة لإجراء اختبارات الفروض وسيتم ذكر تلك الخطوات مع العلم بأننا قد لا نتطرق لشرح تلك الخطوات جميعها، والخطوات هي :

١- صياغة فرض العدم والفرض البديل.

٢- تحديد توزيع المعاينة للاختبار ، مثل توزيع Z ، t ، ... إلخ.

٣- اختيار قيمه α وتحديد القيمة الحرجة للاختبار وقاعدة القرار.

٤- حساب قيمة إحصاءة الاختبار باستخدام بيانات العينة المتاحة.

٥- اتخاذ القرار المناسب.

اختبار المتوسط μ لعينة واحدة بمعلومية الانحراف المعياري σ

One Sample Tests for μ with σ Known

إذا كان الانحراف المعياري للمجتمع معلوم فإنه يمكننا اختبار فرض العدم $H_0 : \mu = \mu_0$ مقابل الفرض البديل والذي قد يكون طرف واحد ($H_a : \mu < \mu_0$ أو $H_a : \mu > \mu_0$) أو طرفين ($H_a : \mu \neq \mu_0$). وتوزيع المعاينة لهذا الاختبار هو التوزيع الطبيعي وإحصاءة الاختبار هي قيمة Z المحسوبة وكما في المعادلة (٨،٣) فإننا نقارن قيمة Z المحسوبة، مع Z الجدولية من الجدول رقم (٧) بالملحق:

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \qquad (8.3)$$

والتي تحدد قيمة الاحتمال لـ α في الطرف المناسب للتوزيع الطبيعي لـ $\overline{X}$.

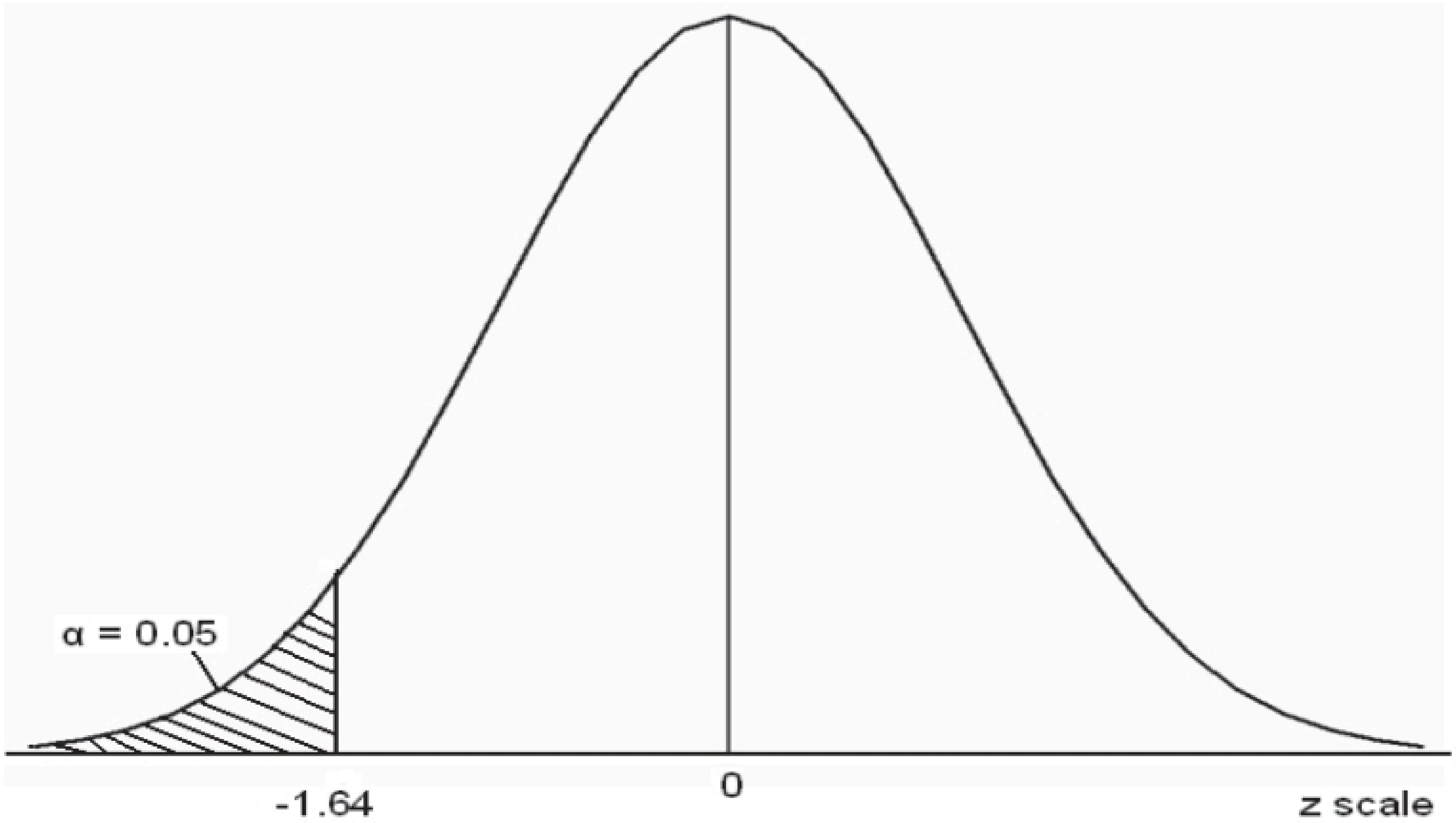
والسؤال كيف نحدد الطرف المناسب، والجواب يتم ذلك بالنظر إلى صياغة الفرض البديل. إذا كان $\mu < \mu_0$ ، فإن لدينا اختبار طرف واحد لليسار وأن الطرف الأيسر للتوزيع يحتوي على α وتكون قيمة Z المحسوبة سالبة وكذلك قيمة Z الجدوليه المستخرجة من الجدول رقم (٧) بالملحق سالبة أيضاً. وبطريقة مشابهة فإذا تمت صياغة الفرض البديل H_a بحيث $\mu > \mu_0$ فإن لدينا اختبار طرف واحد لليمين وتكون المنطقة الحرجة في الطرف الأيمن لتوزيع المعاينة وتكون كلا القيمتين لـ Z

المحسوبة والجدوليه موجبة. أو قد يكون لدينا اختبار الطرفين ناتج من صياغة الفرض البديل حيث $\mu \neq \mu_0$ والذي يتطلب في هذه الحالة تجزئة α إلى قسمين متساويين ووضع كل قسم في كل طرف من أطراف التوزيع ولذلك يكون لدينا منطقتين حرجتين وكذلك قيمتين جدوليتين لـ Z . و Z المحسوبة قد تكون موجبة أو سالبة ولكننا نقارنها بقيمة Z الجدولية لاتخاذ القرار برفض أو عدم رفض فرض العدم.

ولتوضيح النقاش السابق سيتم عرض مثالين أحدهما لاختبار الطرف الواحد والآخر لاختبار الطرفين. وسنبدأ أولاً باختبار الطرف الواحد حيث نفترض بأن الأجر اليومي للعمال الزراعيين يتوزع طبيعياً بمتوسط يساوي ٥٥ دولار (μ =55) وانحراف معياري يساوي ١٠ دولارات ($\sigma = 10$). وقد أدعى العمال في منطقه محددة بأن الأجر اليومي المدفوع لهم زهيد أو قليل. تم أخذ عينه عشوائية حجمها ٣٦ (n=36) من تلك المنطقة وحساب متوسط الأجر لهم فوجد أنه يساوي ٥٢ دولاراً $\overline{X} = 52$. فإذا اخترنا $\alpha = 0.05$ فهل المزارعين في هذه المنطقة يدفعون أجر زهيد للعمال ؟ لحل هذه المسألة فإننا أولاً نصيغ الفرض كالتالي :

$$H_o : \mu = 55$$
$$H_a : \mu < 55$$

حيث إن الفرض البديل عبارة عن اختبار طرف واحد من اليسار. وحيث إن $\alpha = 0.05$ فإن قيمة Z الجدوليه من الجدول (٧) بالملحق تساوي -١,٦٤ . ولذلك فإن المنطقة الحرجة تقع في الطرف الأيسر للتوزيع ومحددة بقيمه Z الحرجة والتي تساوي -١,٦٤ (الشكل رقم ٨,٢).



الشكل رقم (٨.٢). توزيع المعاينة والمنطقة الحرجة لاختبار الطرف الواحد للمتوسط μ .

فإذا كانت قيمة Z المحسوبة تقع لليسار من −١.٦٤ ، أي في المنطقة الحرجة ، فإننا نرفض فرض العدم H_0 ونختم بالقول إن المزارعين في تلك المنطقة يدفعون أجراً زهيداً للعمال. ولكن إذا كانت قيمة Z المحسوبة تقع بين الصفر و −١.٦٤ فإنه لا يمكننا رفض فرض العدم. وباستخدام المعادلة رقم (8.3) يمكن حساب قيمة Z كالتالي :

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{52 - 55}{\frac{10}{\sqrt{36}}} = -1.80$$

ويمكن اتخاذ القرار بمقارنة قيمة Z المحسوبة والتي تساوي -١.٨٠ بقيمة Z الجدوليه والتي تساوي -١.٦٤ . وحيث إن القيمة - ١.٨٠ تقع في المنطقة الحرجة للاختبار فإننا نرفض فرض العدم ونختم القول بأن الأجر المدفوع من قبل المزارعين للعمال في تلك المنطقة أجر زهيد. ويجب ملاحظة أننا لو اخترنا قيمة صغيره لـ α ، ٠.٠١ مثلاً ، فإننا في هذه الحالة لن نرفض فرض العدم.

الآن سنعرض طريقة اختبار الطرفين. حدّد الوكيل المعتاد المسئول عن اختبار أداء آلة تطبيق المبيدات متوسط نقاط التقييم يساوي ٧٠ ($\mu = 70$) وانحراف معياري يساوي ١٥ ($\sigma = 15$). قام الوكيل بتطوير عرض إعلامي متعدد حديث وإعطائه لمجموعة عشوائية عددها ٤٩ من طالبي الترخيص الجدد لمعرفة هل هذا العرض يؤدي إلى أي اختلاف في متوسط النقاط. وقد وجد أن متوسط النقاط لهذه المجموعة يساوي ٧٢ $\overline{X} = 72$ والمطلوب اختبار ذلك باستخدام مستوى معنوية يساوي ٥٪ ($\alpha = 0.05$).

لحل هذا المثال فإننا نحدد أولاً فرض العدم والفرض البديل كالتالي:

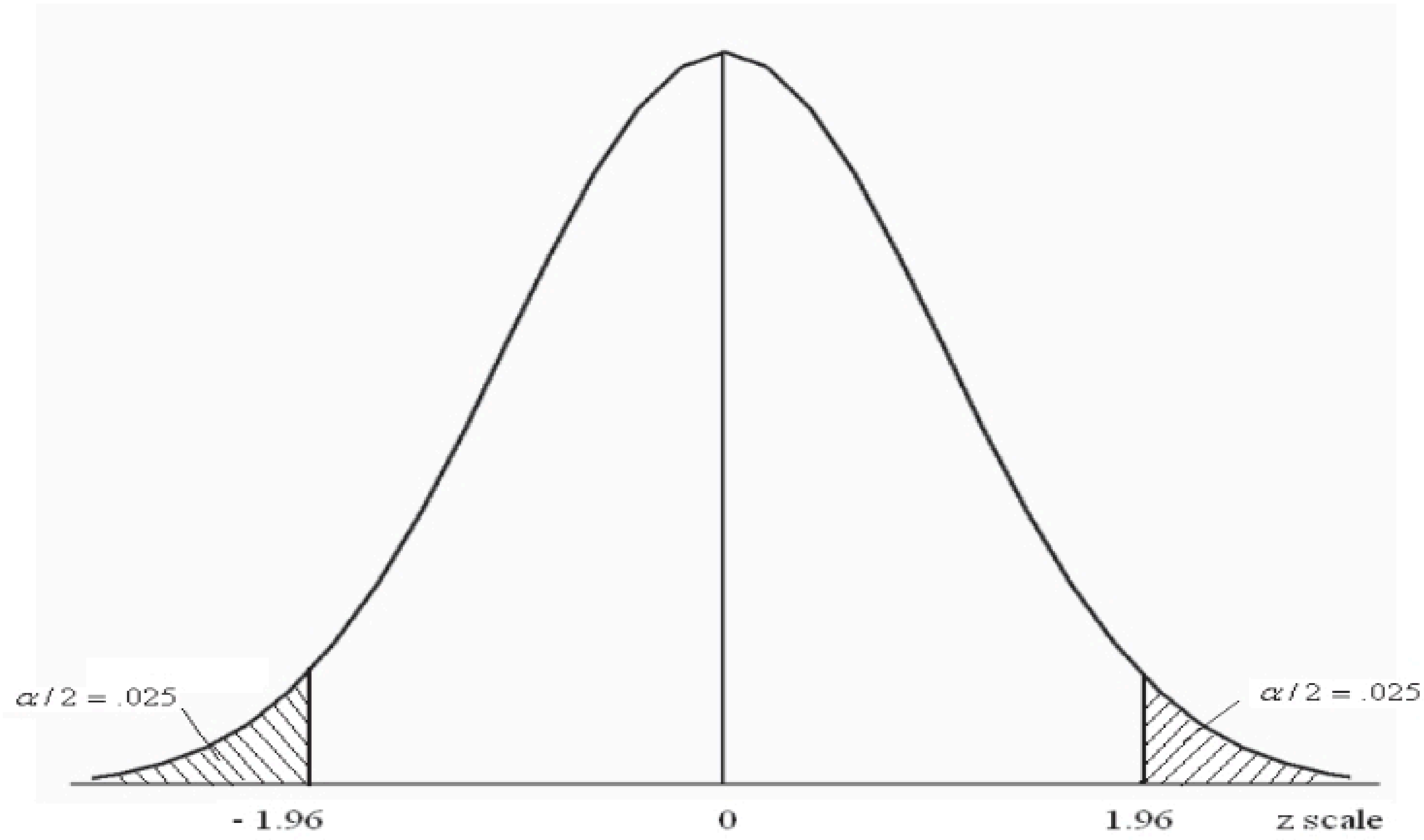
$$H_0 \ : \mu = 70$$

$$H_a \ : \mu \neq 70$$

ثم نقسم α على ٢ ونوجد قيم Z الجدوليه باستخدام الجدول رقم (٧) بالملحق المناظرة لقيم $\alpha/2 = 0.025$ والتي تساوي ١,٩٦ ± ثم نحسب قيم Z كالتالي:

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{72 - 70}{15 / \sqrt{49}} = 0.93$$

لذا عند مقارنة قيمة Z المحسوبة والتي تساوي ٠,٩٣ بقيمة Z الجدوليه ± ١,٩٦ نجد أن قيمة Z المحسوبة تقع بينهما وليست في المنطقة الحرجة (الشكل رقم ٨,٣) ولذا فإننا لا نستطيع رفض فرض العدم. وفي حالة مثالنا ، فإن العرض الإعلامي لن يغيّر متوسط النقاط في اختبار أداة تطبيق المبيدات. ومتوسط العينة الملاحظ كبير ويعود ذلك ببساطة إلى اختلافات المعاينة.



الشكل رقم (٨.٣). توزيع المعاينة والمنطقة الحرجة لاختبار الطرفين للمتوسط μ .

اختبار المتوسط μ لعينة واحدة لها انحراف معياري σ غير معلوم

One Sample Tests for μ with σ Unknown

في هذا الجزء نهدف أيضاً لاختبار الفروض حول المتوسط μ ولكن في هذه الحالة فإننا لا نعرف أي معلومات عن الانحراف المعياري للمجتمع σ. لذا فإننا يجب أن نقدرها باستخدام الانحراف المعياري للعينة S. وتوزيع المعاينة المناسب في هذه الحالة لـ $\overline{X}$ هو توزيع t بدرجة حرية $(n - 1)$.

العينات الكبيرة Large Sample

إذا تم إجراء الاختبار باستخدام عينات كبيرة $(n \geq 30)$ فإنه يمكننا استخدام التوزيع الطبيعي كتقريب لتوزيع t وإجراء الاختبار بطريقة مشابهة لما تم عمله في الجزء السابق. والاختلاف البسيط أن معادلة z المحسوبة تحتوي في المقام على S بدلاً من σ ولذلك فإنه ليس بالضرورة شرح ذلك بمثال.

العينات الصغيرة Small Sample

عند استخدام عينات صغيرة $(n < 30)$ لاختبار الفروض حول المتوسط μ فإن التوزيع المناسب لإجراء هذا الاختبار هو توزيع t. ويمكن توضيح ذلك من خلال المثال التالي:

تحافظ علب الفواكه من الخوخ على متوسط ١٥,٣٣ أوقيه في الحاويات.تم اختيار عينة من ٨ علب وحسب المتوسط للعينة فوجد أنه يساوي ١٥,١٥ أونصه بانحراف معياري يساوي٠,٢ أونصه $(S = 0.2)$. هل بيانات هذه العينة تشير إلى أنه تم المحافظة على متوسط الوزن باستخدام مستوى معنوية يساوي ١٪ $(\alpha = 0.01)$.

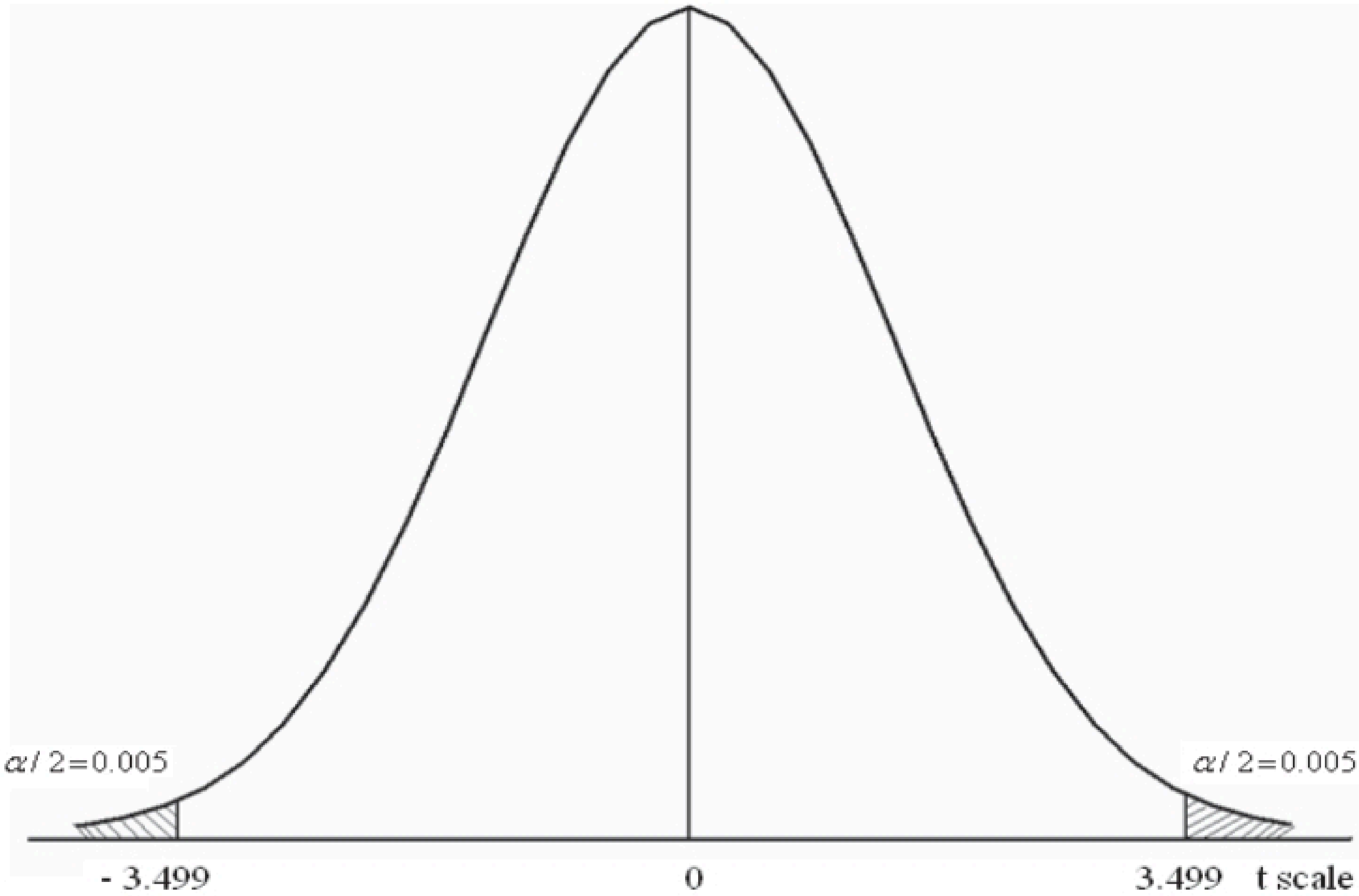
ويمكن صياغة فرض العدم والفرض البديل لهذا الاختبار باستخدام اختبار الطرفين ؛ نظراً لأن اختلاف الوزن يمكن أن يكون في أحد الاتجاهين من الوزن المعياري للخوخ:

$$H_0 : \mu = 15.33$$
$$H_a : \mu \neq 15.33$$

باستخدام توزيع t (الجدول رقم ٨ بالملحق) وعند مستوى معنوية ١٪ ودرجه حرية ٧ $v = (n-1) = (8-1) = 7$ نجد أن قيم t للطرفين هي ±٣,٤٩٩ ، ومن ثمَّ يمكن عرض المناطق الحرجة للاختبار كما في الشكل رقم (٨,٤).

وباستخدام الصيغة الرياضية لـ t يمكننا حساب إحصاءة الاختبار كالتالي:

$$t = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{15.15 - 15.33}{\frac{0.2}{\sqrt{8}}} = -2.54$$



الشكل رقم (٨.٤). المناطق الحرجة لاختبار الطرفين لتوزيع t للمتوسط μ .

وبمقارنة القيمة المحسوبة (−٢,٥٤) بالقيم الجدوليه لـ t نجد أنها لا تقع في المنطقة الحرجة. وعليه فإنه لا يمكننا رفض فرض العدم. وعليه فإن متوسط وزن العينة لا يختلف من الوزن المعياري ١٥,٣٣ أوقيه ماعدا في اختلافات العينة ولذلك فإن المنتج قد وضع الأوزان الصحيحة للخوخ في الحاويات.

اختبار النسبة *P* لعينة واحدة One-Sample Test for *P*

عندما نرغب في إجراء اختبار الفروض لمعلمه التوزيع ذي الحدين *P* ، فإننا نستخدم تقريب التوزيع الطبيعي لتوزيع ذي الحدين ونجري الاختبار بطريقة مشابهة لاختبار سلامة حجر النرد كما تم شرحه سابقاً في جزء الخطأ الأول والخطأ الثاني.

اختبار المتوسطات لعينتين لهما انحرافات معيارية معلومة $\sigma_1 \ , \sigma_2$

Two-Sample Tests for Means with $\sigma_1 \ , \sigma_2$ Known

عندما تكون العينات المختارة من مجتمعين لهما انحراف معياري معلوم فإنه يتم إجراء الاختبار باستخدام التوزيع الطبيعي للمتغير العشوائي $(\overline{X}_1 - \overline{X}_2)$. وتتم صياغة الاختبار بطريقه مماثلة لما تم عمله في حالة العينة الواحدة ما عدا احتواء الفرض على عبارة تمثل الفرق بين متوسطات المجتمع ويتم تعديل الصيغة الرياضية لـ Z المحسوبة لتحتوي على فرق المتوسطات والتي توضحها المعادلة رقم (8.4) التالية :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}}} \qquad (8.4)$$

وغالباً فإن فروض العدم لمعظم الاختبارات في حالة العينتين تصاغ بحيث تعبر عن عدم وجود فرق في متوسطات المجتمع أي $\mu_1 - \mu_2 = 0$ ، ولذا فإنه يتم استخدام الصفر عموماً في حالة استخدام صيغه Z المحسوبة لحساب الفروق. ويمكن عرض المثال التالي لإيضاح كيفية استخدام هذه الطريقة.

أحد علماء محطات التجارب يعرف أن الانحراف المعياري لاستجابة القمح لتطبيق أحد حبيبات الأسمدة المعيارية في المنطقة يساوي ٠,٦ بوشل، وأن الانحراف المعياري لأي تطبيق إضافي يساوي ٠,٥ بوشل. وقد قام العالم باختبار نوع من القمح لتحديد استجابته لنوعي السماد وذلك بتقسيم حوض التجربة لخمسين مربعاً ومعاملة نصف كل مربع بالسماد المعياري والنصف الآخر بالسماد الإضافي. وفي نهاية الموسم كان متوسط إنتاجية القمح الذي تم تسميده بالسماد المعياري تساوي ٤٠ بوشل و٤٢ بوشل للسماد الإضافي. المطلوب باستخدام مستوى ٥٪ $(\alpha = 0.05)$ اختبر هل

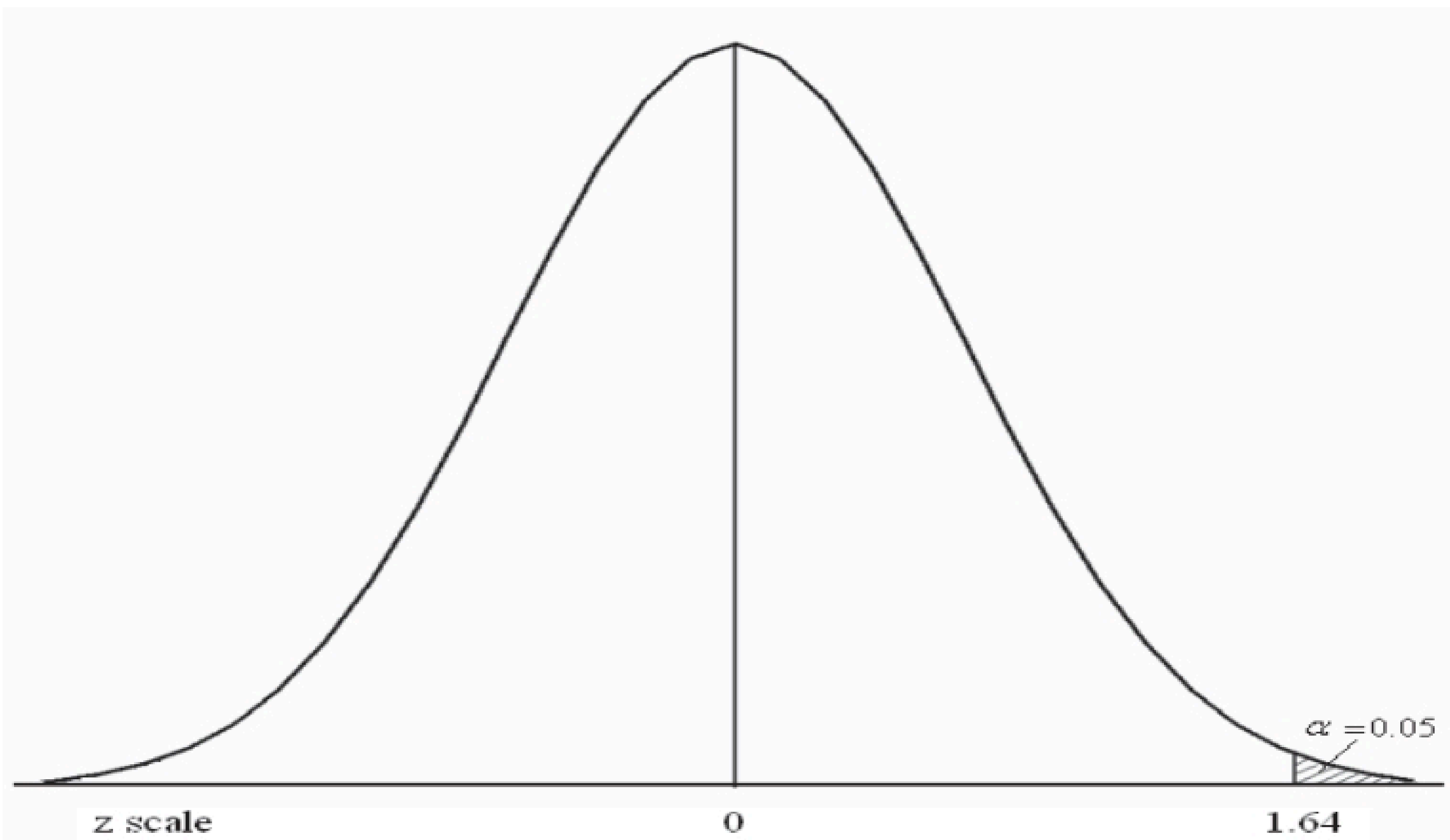
استخدام السماد الإضافي يزيد من متوسط الإنتاجية.

لحل هذه المسألة فإننا أولاً نصيغ الفرض الإحصائي كالتالي :

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 > 0$$

ويمكن عمل هذا الاختبار من طرف واحد باتجاه اليمين باختيار العينة للسماد الإضافي كعينة واحدة. في هذه الحالة فإن القيمة الحرجة لـ Z (من الجدول رقم ٧ بالملحق) تساوي ١,٦٤ (الشكل رقم ٨,٥).

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}}} = \frac{(42 - 40) - 0}{\sqrt{\frac{0.5^2}{25} + \frac{0.6^2}{25}}} = 12.8$$



الشكل رقم (٨.٥). المنطقة الحرجة لاختبار العينتين للمتوسط.

وحيث إن قيمة Z المحسوبة كبيره مقارنة بـ Z الجدولية وتقع خارج المنطقة الحرجة للاختبار، فإننا نرفض فرض العدم ونختم بالقول إن إنتاجية القمح لهذه النوعية من القمح تكون أكبر باستخدام السماد الإضافي .

اختبارات المتوسطات لعينتين لهما انحرافات معيارية غير معلومة σ_1 , σ_2

Two-Sample Tests for Means with σ_1 , σ_2 Unknown

في حالة عدم معرفة الانحراف المعياري لمجتمعات الدراسة فإن توزيع t هو التوزيع المناسب لدراسة الفرق بين متوسطات العينات. ولكن في حالة العينات الكبيرة يمكن استخدام التوزيع الطبيعي كتقريب لتوزيع t واستخدام Z المحسوبة كإحصاء للاختبار. والصيغة الرياضية المستخدمة لحساب قيمه Z هي نفسها تقريباً الصيغة الموضحة في الجزء السابق ماعدا بعض الاختلاف البسيط حيث استبدلت σ_1^2 و σ_2^2 بتباين العينة S_1^2 و S_2^2 ولذلك تكون الصيغة كالتالي :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2{}_1}{n_1} + \frac{S^2{}_2}{n_2}}} \qquad (8.5)$$

ويتم إجراء الاختبار بطريقه مشابهة لما تم عمله في الجزء السابق.

العينات الصغيرة المستقلة Small, Independent Samples

عندما تكون العينات صغيره أقل من ٣٠ $(n_1 , n_2 < 30)$ والمجتمعات مستقلة فإنه يتم إجراء اختبار الفروض باستخدام توزيع t بدرجه حرية $(n_1 + n_2 - 2)$ وحساب قيمه t باستخدام المعادلة رقم (8.6).

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} \quad \text{حيث} \quad S^2 = \frac{(n_1 - 1) S_1^{\ 2} + (n_2 - 1) S_2^{\ 2}}{n_1 + n_2 - 2} \qquad (8.6)$$

والمثال التالي يوضح ذلك، نفترض أن حظيرتين من العجول تمت تغذيتها بعلائق مختلفة لفترة من الزمن. تم إعطاء العجول والبالغ عددها ١٢ في الحظيرة الأولى عليقه بها بروتين مضاف جديد بينما العجول في الحظيرة الثانية والبالغ عددها ٧ عجول أعطيت العليقة العادية دون إضافة كما في الجدول رقم (٨,٢).

الجدول رقم (٨.٢). البيانات لاختبار علائق العجول.

البيان	العليقه المضاف لها البروتين	العليقه العادية
متوسط الوزن المكتسب $(\overline{X})$ رطل	١٢٠	١٠١
$\sum(X - \overline{X})^2$	٥٠٣٢	٢٥٥٢

والسؤال، هل العليقه التي تمت إضافة البروتين لها أكثر فعالية من العليقه العادية عند مستوى معنوية ٥٪.

ولإجراء هذا الاختبار يتم أولاً صياغة فرض العدم والفرض البديل كالتالي:

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 > 0$$

ونحسب درجات الحرية والتي تساوي $v = (n_1 + n_2 - 2) = (12 + 7 - 2) = 17$.

وحيث إن هذا اختبار من طرف واحد باتجاه اليمين نجد من جدول t عند مستوى معنوية ٠,٠٥ (الجدول رقم ٨ بالملحق) أن قيمة t تساوي ١,٧٤. يمكن استخدام

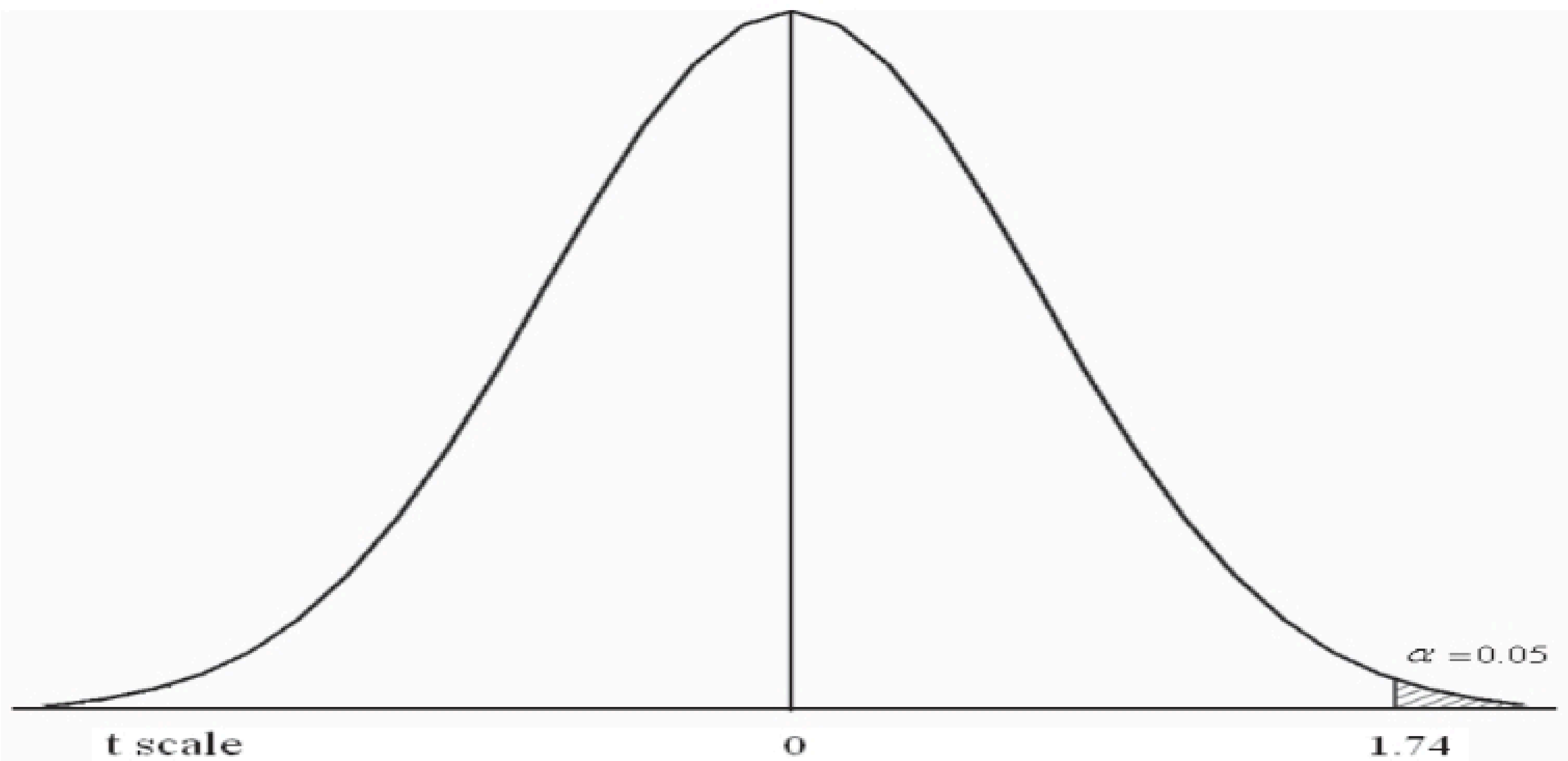
معادلة التباين $S^2 = \sum(X - \overline{X})^2 / n - 1$ والتي ويمكن إعادة صياغتها لتكون $\sum(X - \overline{X})^2 = (n - 1)\, S^2$. ولذلك يمكن إضافة هذه المجموعة (مجموع المربعات) وقسمة مجموعهما على درجة الحرية للحصول على تقدير للتباين المشترك المطلوب لحساب قيمة t .

$$S^2 = (5032 + 2552) / 17 = 446.12$$

وبالتعويض عن هذه القيمة والقيم الأخرى المحسوبة في معادله حساب t يمكن الحصول على قيمة t كالتالي :

$$t = \frac{(120 - 101) - 0}{\sqrt{\frac{446.12}{12} + \frac{446.12}{7}}} = 1.89$$

وحيث إن قيمة t المحسوبة تقع في المنطقة الحرجة للاختبار، فإن القرار يتمثل في رفض فرض العدم. وهذا يعني أن العليقة المضاف لها بروتين جديد تزيد من متوسط وزن العجول المكتسب مقارنة بالعليقة العادية (الشكل رقم ٨،٦).



الشكل رقم (٨.٦). المنطقة الحرجة لاختبار علائق العجول.

العينات الصغيرة المزدوجة Small, Paired Samples

عندما تكون العينات صغيره وغير مستقلة ، فإننا نستخدم اختبار t المزدوج لاختبار الفروض بين متوسطين. درجة الحرية هي عدد الازواج مطروحاً منها الواحد (n - 1) وإحصاءة العينة هي متوسط الفروق بين العينات كما في المعادلة رقم (8.7).

$$\bar{d} = \sum d/n \qquad (8.7)$$

ويمكن إيضاح ذلك من خلال المثال التالي :

تمت تغذية مجموعة أزواج من صغار الحيوانات باستخدام علائق مختلفة ، كما في الجدول التالي :

الجدول رقم (٨.٦). الوزن المكتسب بالرطل لصغار الحيوانات المغذاة بعليقتين مختلفتين.

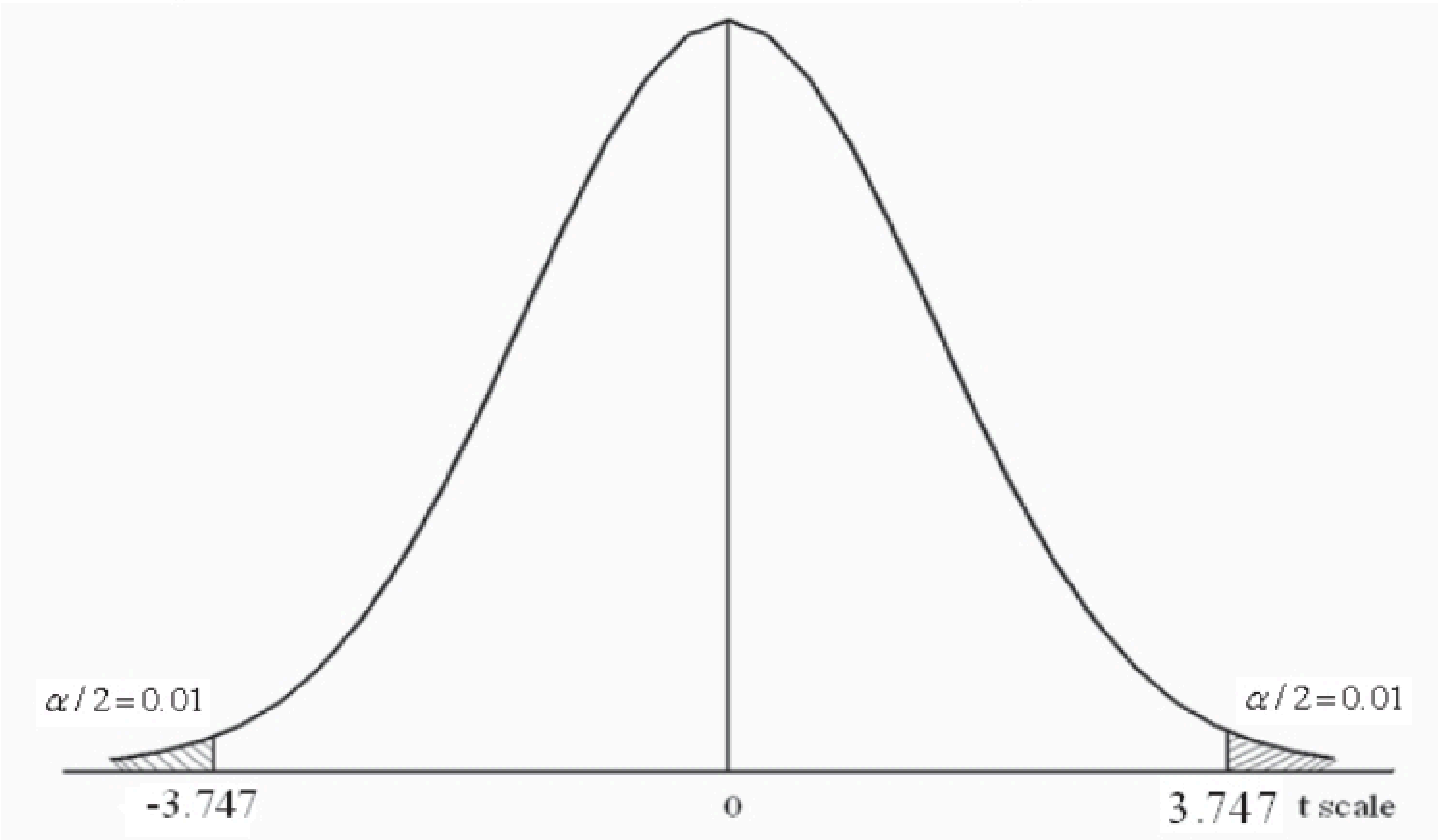
الزوج	عليقه ١	عليقه ٢
١	٢٥	١٩
٢	٣٠	٣٢
٣	٢٨	٢١
٤	٣٤	٣٤
٥	٢٣	١٩

والمطلوب باستخدام مستوى معنوية ٢٪ هل يمكن القول بأن أحد العينتين لها تأثير على الحيوانات يختلف عن الأخرى؟! لإجراء هذا الاختبار فإننا نصيغ فرض العدم والفرض البديل كالتالي :

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

وبحساب درجه الحريـة $4 = (1 - 5) = \nu$ وكـذلك بإسـتخدام الجـدول رقـم (٨) بالملحق نجد أن قيمة t لمستوي معنوية ١٪ $(\alpha / 2)$ تساوي ±٣,٧٤٧ (الشكل رقم ٨,٧) .



الشكل رقم (٨.٧). المناطق الحرجة لاختبار t المزدوج.

ويمكن حساب قيمة t حيث إن الفروق بين العينات تساوى (٦ ، -٢ ، ٧ ، ٠ ، ٤) ولذلك فإن $\bar{d} = 3$ وكذلك $S_d = 3.87$. وعليه فإنه بالتعويض في معادلة t يمكننا حساب قيمتها كالتالي :

$$t = \frac{\bar{d} - (\mu_1 - \mu_2)}{\frac{S_d}{\sqrt{n}}} = \frac{3 - 0}{\frac{3.87}{\sqrt{5}}} = 1.73$$

وحيث إن t المحسوبة لا تقع في المنطقة الحرجة للاختبار فإننا لا نستطيع رفض فرض العدم. ولذلك يمكن القول بعدم وجود فرق بين العلائق بالنسبة للوزن المكتسب للحيوانات في هذه التجربة.

اختبارات النسب لعينتين Two-Sample Tests for Proportions

عندما يكون لدينا عينتان ونرغب في إجراء اختبار الفرق بين النسب في المجتمع يمكننا باستخدام اختبار Z والذي يعتبر كتقريب لتوزيع ذي الحدين. حيث تتم صياغة فرض العدم والفرض البديل كما تم عمله في حالة المتوسطات.

فمثلا يتم طرح النسب من بعضها وتساوى بالصفر في حالة فرض العدم ولا تساوى بالصفر في حالة الفرض البديل وقيمة Z المحسوبة هي عبارة عن الفرق بين نسب العينتين مطروحاً منها الفرق في قيم المجتمع في فرض العدم (تساوي صفر) مقسومة على الخطأ المعياري للفرق في النسب. وبحساب قيمة الخطأ المعياري فإننا نستخدم جميع بيانات العينة ؛ نظراً لأنه في حالة كان فرض العدم صحيح ، فإن كلا العينتين تشمل معلومات من نفس المجتمع. ولذلك يمكن حساب تقديرنا لنسبة المجتمع باستخدام المعادلة رقم (8.8).

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} \tag{8.8}$$

واستخدامها في حساب الخطأ المعياري للفرق في المعادلة رقم (8.9) كالتالي :

$$S_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}} \tag{8.9}$$

ويمكن كتابة Z المحسوبة كما في المعادلة التالية رقم (8.10) :

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{S_{\hat{p}_1 - \hat{p}_2}} \tag{8.10}$$

بينما يتم حساب قيم Z الحرجة باستخدام الجدول رقم (٧) بالملحق. ولتوضيح ذلك يمكن عرض المثال التالي :

يرغب أحد الباحثين في التسويق الذي يعمل في أحد شركات الأغذية في معرفة مدى تأثير الحملة الدعائية على تغيّر نسبة المستهلكين حول تفضيل عصير البرتقال. لذا قام بأخذ عينة مكونة من ٤٠٠ شخص في المدينة التي أقيمت بها الحملة الإعلانية فوجد أن ١٩٢ شخصاً يفضلون عصير البرتقال. وقام بأخذ عينة أخرى حجمها ٥٠٠ شخص من مدينة أخرى والتي لم تقام فيها الحملة فوجد أن ٢١٠ أشخاص يفضلون عصير البرتقال. والمطلوب إجراء اختبار لمعرفة مدى تأثير الحملة الإعلانية على تفضيل المستهلكين لعصير البرتقال باستخدام مستوى معنوية ٥٪.

يتم أولاً صياغة الفرض الإحصائي كالتالي :

$$H_0 : P_1 - P_2 = 0$$

$$H_a : P_1 - P_2 > 0$$

وحيث إن هذا الاختبار هو من طرف واحد فإن قيمة Z الجدوليه (الجدول رقم ٧ بالملحق) تساوي ١٫٦٤. ويمكن حساب تقدير نسبة المجتمع باستخدام المعادلة رقم (8.8) كالتالي : $\hat{P} = (192 + 210) / (400 + 500) = 0.447$. وبذلك يمكننا حساب الخطأ المعياري للفرق في النسبة باستخدام المعادلة رقم (8.7) كالتالي :

$$S_{\hat{P}_1 - \hat{P}_2} = \sqrt{(0.447)(0.553) / 400 + (0.447)(0.553) / 500} = 0.0333$$

وعليه فإن قيمة Z المحسوبة ستكون $Z = [(0.48 - 0.42) - 0] / 0.0333 = 1.80$

وحيث إن Z المحسوبة أكبر من الجدوليه والتي تساوي ١٫٦٤ فإننا نرفض فرض

العدم ونختم القول بأن نسبة المستهلكين الذي يفضلون العصير أعلى في المدينة التي تمت إقامة الحملة الدعائية بها ، وعليه فإن الإعلان مؤثر على الاستهلاك.

تحديد حجم العينة في المسح Determining Sample Size in Surveys

تمت مناقشة مميزات أخذ العينات واستخدامها كقاعدة للاستدلال الإحصائي. ولكن في معظم الحالات لا نعلم الحجم الأفضل للعينة المختارة. وفي مثل هذه الحالات فإننا نحدد قيمة الخطأ الممكن قبوله ومن ثم يتم حساب حجم العينة المناسب بناء على ذلك.

الاستدلال الإحصائي للمتوسط μ For Statistical inference on

عندما يكون الهدف هو الاستدلال حول متوسط المجتمع ، μ ، فإن قيمة الخطأ المتوقع معطى بالفرق $(\overline{X} - \mu)$ والذي يمثل بعد المتوسط المقدر $\overline{X}$ عن متوسط المجتمع μ ؛ نتيجة اختلافات العينة. فلو تم تحديد هذا الخطأ بقيمة معينة ولنقل E ، فإنه في هذه الحالة يمكننا معالجة الصيغة الرياضية لـ Z وذلك لتحديد حجم العينة الضروري والذي يمكن معه الوصول لهذا المقدار من الخطأ ، وكذلك الحال يمكن تحديد مستوي فترة الثقة للاستدلال مثلا ٩٩٪ ، ٩٥٪ ، ... إلخ. ونحن نتذكر بأنه يمكن حساب قيمه Z كما في المعادلة رقم (8.11)) التالية :

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad (8.11)$$

فإذا كان مستوى الثقة المطلوب هو $(1 - \alpha)$ فإنه يمكننا كتابة فترة الثقة ، كما في المعادلة التالية رقم (8.12).

$$-Z_{\alpha/2} \cdot \sigma / \sqrt{n} \leq \overline{X} - \mu \leq Z_{\alpha/2} \cdot \sigma / \sqrt{n} \qquad (8.12)$$

وكون التوزيع الطبيعي توزيع متماثل فإنه يمكننا التركيز فقط على حالة اللامساوة في الطرف الأيمن والتي يمكن التعبير عنها بالمعادلة رقم (8.13) التالية :

$$\overline{X} - \mu \leq Z_{\alpha/2} . \sigma / \sqrt{n} \qquad (8.13)$$

والتي تتضمن أن أكبر قيمه يمكن افتراضها لـ $\overline{X} - \mu$ هي $Z_{\alpha/2} . \sigma/\sqrt{n}$ ولكن قد تم التعبير سابقاً عن أكبر خطأ نرغبه بالرمز $E = \overline{X} - \mu$. لذا يمكن إعادة كتابة المعادلة أعلاه للتعبير عن الخطأ المسموح به بالمعادلة رقم (8.14) التالية :

$$E = Z_{\propto/2} . \sigma / \sqrt{n} \qquad (8.14)$$

وبحل المعادلة رقم (8.14) يمكن إيجاد قيمه n وحجم العينة اللازم، والذي يضمن بمستوى معنوية $100(1-\alpha)$ أن $\overline{X} - \mu$ لن يزيد عن خطأ المعاينة E. وعليه فإن المعادلة التي تحدد حجم العينة الأمثل يمكن كتابتها بالشكل التالي (8.15).

$$n = \frac{Z^2{}_{\alpha/2} \, \sigma^2}{E^2} \qquad (8.15)$$

ويمكننا تأمل المثال التالي لإيضاح ذلك. إحدى شركات البيوت المحمية والتي تستخدم هياكل للبيوت وتغطيها بالبلاستيك الشفاف ترغب في شراء تلك الهياكل من أحد شركات البلاستيك الجديدة والتي تقع بالقرب من مصنعها. وتعلم هذه الشركة من واقع خبرتها السابقة أن الانحراف المعياري لأطوال تلك الهياكل يساوي ٠,٢٠ قدم. المطلوب تحديد حجم العينة اللازم لتقدير متوسط الطول لمجتمع الهياكل بخطأ مسموح يساوي ٠,٠٣ وبدرجة ثقة٩٥٪ للتقدير؟

لحل هذه المسألة فإن المعطيات هي $\sigma = 0.2$ ، $E = 0.03$ و $Z_{\alpha/2} = 1.96$

وبالتعويض عن هذه القيم في معادلة حساب حجم العينة الأمثل n نحصل على:

$$n = \frac{Z^2_{\propto/2}\sigma^2}{E^2} = \frac{1.96^2\,(0.2^2)}{0.03^2} = 170.7$$

وعليه فإن شركة البيوت المحمية يجب أن تأخذ عينة حجمها ١٧١ هيكلاً من الشركة الجديدة لتقدير متوسط الطول بخطأ ٠,٠٣ ومستوى ثقة ٩٥٪.

الاستدلال الإحصائي لـ P For Statistical Inference on

يمكن تحديد حجم العينة لمعلمة التوزيع ذي الحدين P إذا تم تحديد أكبر حجم للخطأ المسموح به مقدماً، $E = \hat{P} - P$ وكذلك مستوى المعنوية للاختبار. وفي أمثلة عديدة، فإنه لا يوجد قيمة لـ $\hat{P}$ يمكن استخدامها لحساب σ_P؛ نظراً لأن العينة لم تسحب حتى الآن. وفي مثل هذه الحالات فإننا نستخدم التقدير $P = \frac{1}{2}$؛ نظراً لأنه يمكننا إثبات أن قيمه σ_P ستصل إلى أكبر قيمة لها عندما تكون $P = \frac{1}{2}$. وعليه فإنه في حالة التعويض عن $P = \frac{1}{2}$ في معادله Z حيث $Z = (\hat{P} - P) / \sqrt{Pq/n}$ لنحصل على $Z = (\hat{P} - P) / \sqrt{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)/n}$ وحل المعادلة لإيجاد قيمة n يمكن الوصول للمعادلة المطلوبة (8.16):

$$n = (p)(1-p)\frac{Z^2_{\alpha/2}}{E^2} = \frac{Z^2_{\alpha/2}}{4E^2} \qquad (8.16)$$

ويمكن توضيح ذلك من خلال المثال التالي.

نفترض أن شركة البطيخ والتي تقوم بقطف المحصول ترغب في الحصول على

تقدير لنسبة البطيخ التالف والذي يذهب للمصنع بدرجه دقة ٣٪ ومستوى ثقة يساوي ٩٥٪ فكم حجم العينة الواجب أخذه.

في هذه الحالة فإن $E=0.03$ وكذلك $Z_{\alpha/2}=1.96$ ولذلك وبالتعويض في المعادلة آنفاً فإن $n=1.96^2/(4)\,(0.03)^2=1067.1$ أو ١٠٦٨ بطيخة. ويمكن تقليل حجم العينة بطريقة ما إذا كان لدينا تقدير لـ P والتي تعبّر عن نسبة التالف من فترة زمنيه سابقه، ... إلخ.

حجم العينة اللازم للتحكم في α و β

Sample Size to control Both α and β

في اختبارات الفروض نرغب في الغالب في تحديد مستوي α عند مستوى معين مثلاً $\alpha=0.05$ وكذلك نرغب أيضاً في تحديد قيمه β لقيمة معطاة للفرض البديل H_a. ويمكننا عمل ذلك شريطة حساب حجم العينة الذي يمكننا من التحكم في الأخطاء عند مستوياتها المرغوبة واستخدام هذا الحجم عند إجراء الاختبار.

وعند صياغتنا للفروض سابقاً، تم اختيار قاعدة القرار بناء على القيم الحرجة لـ Z والتي تحدد احتمالات α في أحد أطراف التوزيع الاحتمالي للفرض الإحصائي. وبعد تحديد قاعدة القرار يتم حساب قيمه β كنسبة من توزيع المعاينة للفرض البديل والتي تتقاطع مع خط قاعدة القرار في منطقه القبول لفرض العدم. لذا فإنه يمكن التعبير عن القيم الحرجة لقاعدة القرار بصيغتين (معادلة 8.17 و 8.18) إحداهما مبنية على توزيع المعاينة عندما تكون H_0 صحيحة والأخرى مبنية على توزيع المعاينة عندما يكون الفرض البديل H_a صحيح. وكلا الصيغتين تحتوي على حجم العينة n والتي يمكن حل تلك المعادلة عند تحديد قيمة β بقيمه محددة مسبقاً.

$$\overline{X}_c = \mu_0 - Z_\alpha . \sigma / \sqrt{n} \qquad (8.17)$$

$$\overline{X}_c = \mu_a + Z_\beta . \sigma / \sqrt{n} \qquad (8.18)$$

وكلا الصياغتين آنفاً تعبّر عن اختبار من طرف واحد (الأيسر). ولإجراء الاختبار باتجاه الطرف الأيمن يمكن عكس الإشارة الموجبة والسالبة في كلا التعبيرين بحيث تحتوي المعادلة الأولى على إشارة موجبة والمعادلة الثانية على إشارة سالبة ولإجراء اختبار الطرفين يتم استبدال Z_α بـ $Z_{\alpha/2}$ في المعادلة (8.17). ونظراً لأن كلا الصياغتين تحتوي على $\overline{X}_c$ يمكن مساواتهم ببعض وحلهما لإيجاد n التي نرغب في إيجادها (معادلة رقم 8.19).

$$n = \left[\frac{(Z_\alpha + Z_\beta)\sigma}{(\mu_0 - \mu_a)} \right]^2 \qquad (8.19)$$

ويجب ملاحظة أنه يتم استخدام القيم الموجبة لـ Z في المعادلة آنفاً رقم (8.19)؛ لأن الإشارات في المعادلات الخاصة بـ $\overline{X}_c$ تؤدي للحصول على القيم الصحيحة.

ويمكن إيضاح ذلك من خلال المثال التالي.

ترغب شركة أعمال زراعيه كبرى في دراسة السوق لمعرفة الحاجة لخريجي البكالوريوس في إدارة الأعمال الجدد. وتعتقد المؤسسة بأنه ليس هناك تغيراً في السوق وترغب في اختبار فرض العدم القائل بأن متوسط الراتب للخريجين الجدد يساوي ٢٢٥٠ دولاراً في الشهر بمستوى معنوية ٥٪ وأيضاً ترغب أن يكون الفرق مساوياً لـ١٠٠ دولار شهرياً باحتمال $\beta = 0.03$. فإذا كان الانحراف المعياري لمجتمع الراتب للمعينين الجدد يساوي $\sigma = 300$ دولار، فكم حجم العينة الواجب أخذه؟

حيث إن هذا الاختبار عبارة عن اختبار طرفين والذي يمكن صياغة الفرض

كالتالي :

$$H_0 : \mu = 2250$$
$$H_a : \mu \neq 2250$$

لذا فإن α تقسم على ٢ وباستخدام جدول Z رقم (٧) بالملحق نجد أن قيمة Z تساوي١,٩٦ وحيث إن β تساوي ٠,٠٣ فإن Z_β تساوي ١,٨٨من نفس الجدول . ونهدف لمعرفة هل الراتب الابتدائي للخريجين يتغير بمعدل ١٠٠دولار شهرياً، لذا فإن القيمة المرتبطة بالفرض البديل تختلف بـ ١٠٠دولار أي، $100 = (\mu_0 - \mu_a)$. والآن يمكن التعويض بهذه القيم في الصيغة الرياضية ل n لإيجاد الحجم المناسب للعينة كالتالي :

$$n = \left[\frac{(Z_\alpha + Z_\beta)\sigma}{(\mu_0 - \mu_a)}\right]^2 = \left[\frac{(1.96 + 1.88)\,(300)}{(100)}\right]^2 = 132.7$$

وعليه فإن الشركة تحتاج لعينه حجمها ١٣٣ موظفاً جديداً من خريجي إدارة الأعمال المزرعية لإجراء الدراسة.

ملحق للفصل الثامن Appendix to Chapter 8

استخدام برنامج اكسل لإجراء اختبار العينتين للفروض
Using Excel to perform Two-Sample tests of Hypotheses

اختبار عينتين باستخدام Z Tow-Sample z test

يمكن إجراء الاختبار السابق باستخدام بعض البرامج الخاصة مثل برنامج اكسل ويتم ذلك بوضع المتوسط الأول في الخلية A1 والمتوسط الثاني في الخلية B1 ثم اختيار الأمر تحليل البيانات من قائمة الأدوات ثم اختيار Z لعينه للمتوسط. يظهر جدول بعد ذلك يتم إكماله بوضح المدى A1 للمتوسط الأول في المستطيل الأعلى والمدى B1 للمتوسط الثاني في المستطيل الثاني. نضع صفر في المستطيل الخاص بفرق المتوسطات، ثم نحسب التباين المعلوم σ^2/n للمتغيرين الأول والثاني ثم نضعهما في المستطيلات الخاصة بها. نغيّر قيمة α إذا لم تكن نفس المستوى المطلوب. وأخيراً نختار المكان الذي نرغب أن نظهر فيه النتيجة وليكن مثلاً D1 ثم نختار التنفيذ. سوف تظهر النتيجة في نفس ورقة العمل كما في الجدول رقم (٨.٤).

الجدول رقم (٨.٤). مخرجات ورقة العمل لاختبار Z للعينتين.

اختبار Z : عينتين للمتوسط

	المتغير الأول	المتغير الثاني
المتوسط	٤٢	٤٠
التباين	٠,٠١	٠,٠١٤٤
عدد المشاهدات	١	١
الوسط الافتراضي	صفر	
الفرق		
Z	١٢,٨٠٣٦٩	
$P(Z \le z)$ طرف واحد	صفر	
قيمة Z الحرجة لطرف واحد	١,٦٤٤٨٥٣	
$P(Z \le z)$ طرفين	صفر	
قيمه Z الحرجة لطرفين	١,٩٥٩٩٦١	

ولقراءة العمود الأول يجب زيادة عرض العمود بحيث يكون ٢٠ أو أكثر. ويجب ملاحظة أن الإجراء يفترض أن البيانات الأولية في الأعمدة A و B من ورقة العمل. ويجب حساب التباين بنفس الطريقة المعتادة . ولو كان هناك قيم أوليه للبيانات في هذه الأعمدة فإنه يمكن وضع قيم التباين لها.

اختبار t للعينات المستقلة Independent t Test

يتطلب هذا الاختبار أن تكون لدينا البيانات الأولية بعكس ملخص البيانات التي تم استخدامها في المثال السابق. وعليه يمكن إيضاح ذلك من خلال المثال التالي .

تم زراعة عدد ٢٠حوضاً تجريبياً بمحصول الذرة وتم تسميد نصف هذه الأحواض باستخدام نوعية سماد جديدة عند المعدل الطبيعي بينما تم تسميد النصف الباقي بالسماد القديم عند نفس المعدل. وقد تم حساب الإنتاجية للايكر وكانت كالتالي :

السماد الجديد	٨٥	٩١	٨٦	١٠٢	٩٤	٩٨	٨٨	١٠٥	١١٢	٨٢
السماد القديم	٨٢	٩٥	٨٠	٩٧	٩١	٩٠	٨٧	١٠٠	١٠٣	٨٠

المطلوب هل يوجد علامات تبيّن أفضلية السماد الجديد باستخدام مستوى معنوية ١٪.

لعمل هذا التحليل باستخدام برنامج الإكسل نقوم بإدخال البيانات في ورقة عمل في البرنامج في العمودين A1 و B1 للمتغيرين مع وضع عنوان لهما ثم نختار من قائمة الأدوات الخيار تحليل البيانات ثم نختار اختبار t لعينتين بافتراض تساوي التباين لهما . يظهر لنا جدول ندخل المدى للمتغير الأول $A_1 : A_{11}$ والمدى للمتغير الثاني

$B_1 : B_{11}$ ونضع الرقم صفر في المستطيل الخاص بفرق المتوسط ثم نختار العلامة أمام العلامات ؛ نظراً لوجود أسماء للمتغيرات في الصف الأول من ورقة العمل ، نغير قيم مستوى المعنوية α إلى ٠,٠١ ثم نختار مكان المخرجات لتكون في الخلية D_1 ثم ننفذ الأمر. تظهر لنا النتائج الموضحة في الجدول رقم (٨,٥).

الجدول رقم (٨.٥). مخرجات ورقة العمل لاختبار t لعينتين

اختبار t : حاله عينتين بافتراض تساوي التباين		
	السماد الجديد	**السماد القديم**
المتوسط	٩٤,٣	٩٠,٥
التباين	٩٥,٣٤٤٤٤٤٤٤	٦٨,٢٧٧٧٧٧٧٨
عدد المشاهدات	١٠	١٠
التباين التجميعي	٨١,٨١١١١١١١	
الوسط الافتراضي	صفر	
الفرق		
درجه الحريه df	١٨	
إحصاءة t	٠,٩٣٩٤٢٥٧٣٥	
P(T <= t) طرف واحد	٠,١٧٩٩٧٤٣١٩	
قيمة t الحرجه لطرف واحد	٢,٥٥٢٣٧٨٦٤٦	
P(T <= t) لطرفين	٠,٣٥٩٩٤٨٦٣٨	
قيمة t الحرجه لطرفين	٢,٨٧٨٤٤١٥٩٢	

ومن الجدول نلاحظ أن قيمة t المحسوبة تساوي ٠,٩٣٩٤ والتي تكون معنوية عند المستوى ٠,١٧ وليس ٠,٠١ والتي تم تحديدها من قبلنا ، لذا فإننا لا نستطيع رفض

فرض العدم. والقيمة الحرجة لـ t تساوي ٢,٥٥ والتي قيمتها أكبر من قيمة t المحسوبة. درجه الحرية لهذا الاختبار تساوي ١٨. متوسط إنتاجيه القمح باستخدام السماد الجديد تساوي ٩٤,٣ بوشل /ايكر بينما المتوسط باستخدام السماد القديم يساوي ٩٠,٥ بوشل/ايكر. التباين التجميعي يساوي ٨١,٨١.

تمارين Exercises

١- ما الفرق بين المعلمة والإحصاءة؟

٢- حلل العبارة التالية: في اختبارات الفروض الإحصائية، يجب أن نحدد دائما مستوى المعنوية أو (α) بقيمة صغيره جداً مثلاً ٠,٠١ ، وعليه يمكننا تجنب الوقوع في كثير من الأخطاء.

٣- المطلوب صياغة فرض العدم والفرض البديل للحالات التالية:

أ) ترغب شركة لصناعة الحبوب في اختبار عملياتها الإنتاجية نظراً لشكوى تقدمت بها منظمه المستهلكين والمتضمنة أنها تضع حبوب وزنها أقل من ١٤ أوقية في صناديق مقاسها ١٤ أوقية.

ب) يدعى مجلس الإدارة للمزارعين بأن نسبة المزارعين المفضلين للقوانين الصارمة للتحكم البيئي متساوية في ولايتي أوكلاهوما ونيوجرسي.

ج) ادعت الشركة الصانعة بأن متوسط العرض لقفازات اللحام السوداء هو ٢ بوصة إلاّ أن صاحب المتجر الزراعي شك في ذلك.

د) يرغب أحد الاقتصاديين في معرفة هل نسبة البطالة في المناطق الريفية أقل منها العام الماضي علماً بأن الاقتصاد الكلي قد حقق نمواً هذه السنة.

هـ) يرغب مدير أحد محلات العرض الزراعية في اختبار الفرض القائل بأن

متوسط رصيد الحساب الائتماني لمؤسسته يساوي على الأقل ٧٠ دولاراً.

٤- انتقد العبارات التالية :

أ) نظراً لأهمية التحكم بالخطأ من النوع الأول مقارنة بالنوع الثاني ، فإنه يجب صياغة اختباراتنا بحيث تكون حساسة لأخطاء النوع الأول فقط.

ب) نظراً لأن تقليل الخطأ من النوع الأول يزيد من أخطاء النوع الثاني ، فإنه من المستحيل التحكم فيهما معاً.

ج) ما دام أننا اتخذنا قرار باختبار الفروض ، فإنه من الممكن وقوع كلا النوعين من الأخطاء النوع الأول والنوع الثاني.

٥- المواصفات الخاصة بمكونات القطن المحصود تتطلب أن لا يقل متوسط الطول عن ١٠٠ بوصة بانحراف معياري ٨ بوصة. تم اختيار عينة عشوائية حجمها ٨١ من شحنه كبيره جداً وثم حساب متوسطها وانحرافها المعياري فوجدت ٩٨.٥ و ٨ بوصه على التوالي.

أ) بناء على تحليلك هل ستقبل تلك الشحنة؟ برّر إجابتك.

ب) بناء على القيمة الحرجة في الجزء (أ) ما هو احتمال الخطأ من النوع الثاني لمكونات الشحنة لمتوسط طول ٩٧ بوصة وانحراف معياري ٨ بوصة.

٦- تم إعطاء اختبار تحديد المهارات القياسي للمزارعين لعدة سنوات لطلاب الثانويات الزراعية. حيث كان المتوسط ٧٥ بانحراف معياري ٧. قرر المعلم زيادة معدلات الطلاب بزيادة التركيز على التعليم. لذا تم إعطاء التعليمات لمعلم الفصل المكون من ٢٥ طالباً بزيادة التركيز على التعليم وزاد المعدل إلى ٧٨ في الاختبار. هل يمكننا القول بمستوى معنوية ٥٪ بأن التعليمات لزيادة التركيز على التعليم أدت لزيادة متوسط درجه الاختبار؟

٧- عند حقن الدواجن بنوع معين من هرمونات النمو يؤدي ذلك لزيادة متوسط وزن الدجاج بـ ٠,٥ رطل. قام خبير التحكم في الجودة في مزارع هاووي بأخذ عينة عشوائية حجمها ٩ دجاجات ثم حقنها بالهرمون وحسب متوسط الوزن المكتسب فوجد أنه يساوي ٠,٦ رطل بانحراف معياري ٠,٢ إذا تم إجراء اختبار من طرفين عند مستوى معنوية ٥٪ فهل ذلك يؤكد أن متوسط زيادة الوزن تساوي ٠,٥رطل.

٨- قام اثنان من المختصين في البساتين بتقييم الاختلافات في النباتات. كان التقييم الأول لعدد ١٢ نبات بمتوسط ٩٠ بينما كان التقييم الثاني لعدد ٨ نباتات بمتوسط ٨٥. تشير الخبرات السابقة أن التباين في التقييم يساوي ٤٠. المطلوب باستخدام مستوى معنوية ٥٪ اختبر مدى وجود اختلاف في تقييم الاثنين من عدمه.

٩- نرغب في اختبار مدى قابلية منتج الفواكه الجديد . فإذا كان نصف المستهلكين الذين ثم اختبارهم قرروا شراء المنتج سيتم تسويقه وإن كان غير ذلك فلن نقوم بتسويقه. تم اختيار عينه عشوائية حجمها ٦٤ مستهلكاً من٢٦ ولاية التي ستقوم بشراء المنتج. اختبر بمستوى معنوية ٥٪ هل يجب تسويق ذلك المنتج أم لا.

١٠- تم اختبار نوعين من المديا السائلة(وسط زراعة مائي) لزراعه الطماطم مائياً (بدون تربه) تمت زراعة عدد ١٠نباتات في النوع الأول والتي أنتجت ٢٨ من الطماطم بانحراف معياري ٢ بينما تمت زراعة ١٥ نباتاً في النوع الثاني من نفس النوعية وأنتجت ١٨ من الطماطم بانحراف معياري يساوي ٩. أختبر هل النوع الأول (وسط الزراعة المائي) أفضل من الثاني عند مستوى معنوية٥٪.

١١- مستثمر زراعي لديه موقعين مفصولة ببضعة أميال وكذلك لديه قطيعين ماشية. في المكان الذي يقيم فيه المستثمر تتم إدارة القطيع بحيث يكون هناك إنتاج من العجول في فصل الخريف وفي المكان الآخر يكون الإنتاج من العجول في فصل الربيع. عدد

الحيوانات في الموقع الذي يقيم فيه ٤٢٠ بقره وقد أنتجت ٣٧٨ عجلاً في الخريف الماضي بينما عدد الأبقار في الموقع الآخر ١١٠بقره وقد أنتجت ٩٠ عجلاً في الربيع الماضي. اختبر بمستوى معنوية ٥٪ هل نسبة العجول في كلا الموقعين متساوية.

١٢- تم إجراء اختبار أولي لعدد ٨ طلاب من الزراعة ،ثم تم إعطاءهم تعليمات لطريقة التطعيم والتصفيف للنبات ثم أعيد اختبارهم مرة أخرى فكانت نتائجهم كالتالي:

الطالب	الدرجة قبل التدريب	الدرجة بعد التدريب
١	٣٧	٧٥
٢	٦٢	٨٧
٣	٧١	٩٥
٤	٤٥	٧٢
٥	٥٣	٧٨
٦	٢٤	٦٣
٧	٩٠	٩٨
٨	٨٦	٨٤

اختبر بمستوى معنوية ١٪ هل للتعليمات المعطاة لهم تأثير على الدرجة المتحصل عليها.

١٣- يرغب أحد الاقتصاديين الزراعيين في تحديد متوسط معدل الأجر لعمّال التبغ في منطقه الزراعة المسماة فلو-كورد في ولاية كارولينا، جورجيا وفلوريدا. وقد قام بدراسة استطلاعيه فوجد أن متوسط الأجر ٦,٨ دولار. فإذا كان يرغب في إيجاد فترة الثقة ٩٥,٥٪ لأقصى خطأ لتقدير للأجر بحيث لا يزيد عن ٠,١ دولار فكم حجم العينة المطلوب لإجراء الدراسة.

١٤- يتوقع وصول شحنة كبيره من الحبوب إلى صالة الحبوب ويجب أخذ عينة منها لتحديد محتوياتها من القش. كم يجب أن يكون حجم العينة اللازم أخذها إذا كنا نرغب أن تكون نسبة القش في العينة في حدود ١,٥٪ بدرجة ثقة ٩٩٪ ؟ وبافتراض أن نسبة محتويات القش لا تزيد عن ٠,١ بناء على دراسات سابقة لهذه الحبوب.

١٥- حل التمرين١٢ باستخدام برنامج اكسل.

١٦- حل التمرين التالي باستخدام برنامج اكسل.

يرغب أحد المزارعين في عزل محله التجاري ورغبة منه في تحديد المواد المستخدمة وجد دراسة بها بيانات عن قوة الأثر لنوعين من المواد كالتالي. ولكن لا يوجد تحليل لتلك البيانات. المطلوب اختبر بمستوى معنوية ٥٪ أي تلك المواد لها تأثير أكبر في العزل.

العازل ١	١,٥	١,٤١	١,٥٨	١,٤	١,٤٥	١,٥٨	١,٥٣	١,٤٦
العازل ٢	١,١٤	١,٢٢	١,٢	١,١٩	١,٢٧	١,٢٤	١,٣١	١,٢٣

الفصل التاسع

تحليل التبايـــن

Analysis of Variance

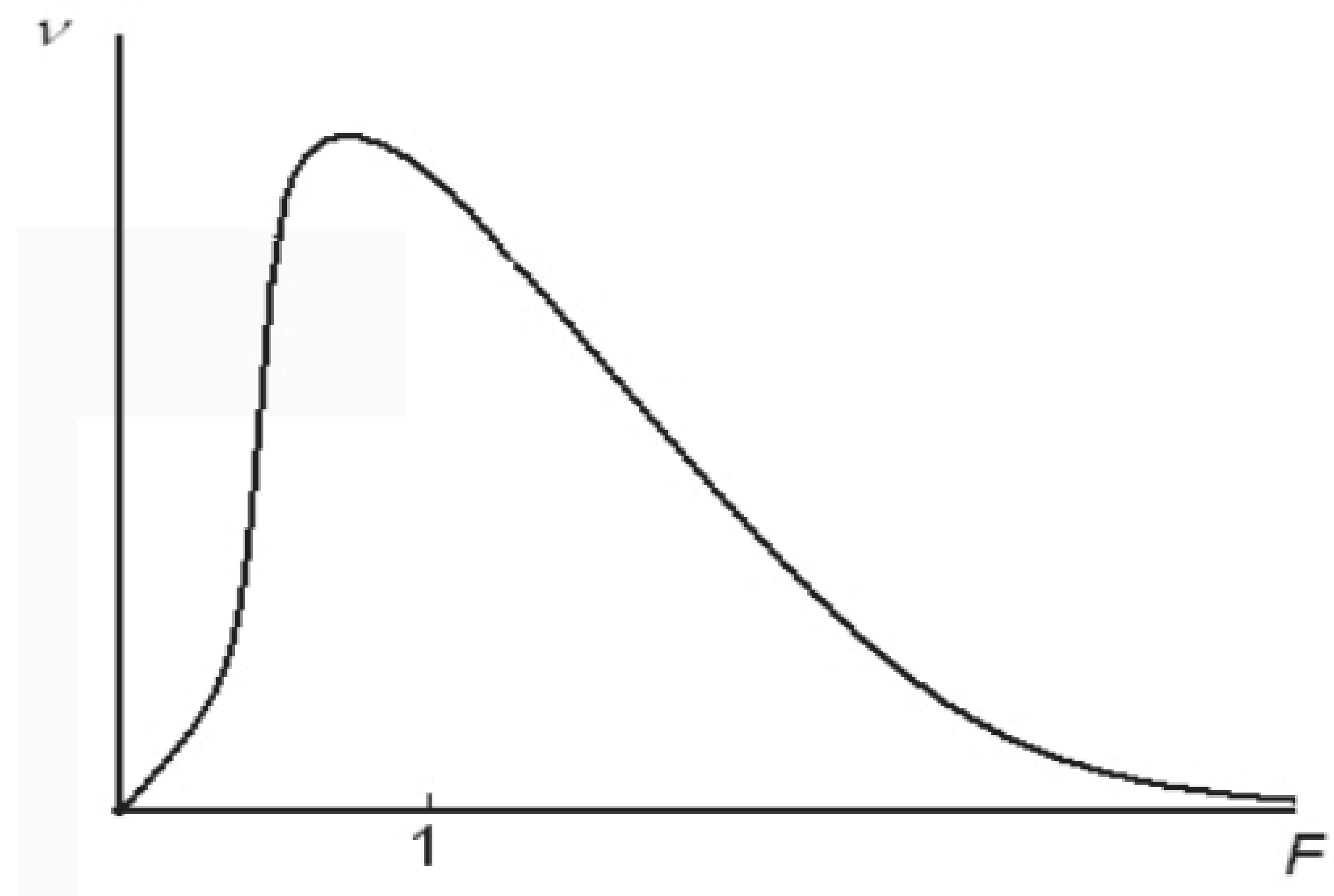
في الفصول السابقة تمت دراسة الاستدلال الإحصائي وقمنا بعرض اختبارات الفروض لقيم المعالم في حالة وجود مجتمع أو مجتمعين. في هذا الفصل سيتم التطرق لطريقة إجراء اختبارات الفروض في حالة وجود ثلاثة مجتمعات أو أكثر. وتكمن أهمية هذا الموضوع ؛ نظراً لأن المقارنات بين الأزواج المتعاقبة عمليه مرهقة وغالبا لا تعطي في النهاية نتائج دقيقه. فمثلاً إذا كان لدينا اختبار لثلاثة متوسطات ، يمكننا اختبار جميع الأزواج الممكنة من المتوسطات باستخدام الطريقة التي تم عرضها في الفصول السابقة ولكن هذا الاختبار لا يمكن الاعتماد عليه ، ويعطي تحليل التباين طريقة جيدة لإجراء الاختبارات لثلاثة متوسطات أو أكثر ذات موثقية عالية.

وحيث إن تحليل التباين له مصطلحات خاصة تختلف نوعاً ما عن السابق فإننا سنتطرق أولاً للحديث عن تلك المصطلحات. تستخدم بيانات العينة لمقارنة مجموعة مستويات من المتغيرات العشوائية تدعى المعاملة أو المعالجة. ويشير مصطلح المعاملة إلى أي متغير عشوائي كضابط للباحث. فمثلاً عالم الحشرات ربما

يرغب في اختبار مدى فعالية مجموعة من المستويات من مبيد حشري في التحكم في فراشة معينة على شجر التفاح كمعاملة. أو عالم الحيوان ربما يختار تركيزات مختلفة من البروتين في تغذية الأبقار كمعاملة أو معالجة. المصطلح الثاني هو وحدة التجربة. وحدة التجربة هي عبارة عن الكيان الذي يتم معاملته. ففي مثال عالم الحشرات وحدة التجربة ربما تكون الشجرة أو قطعة من الأرض ذات مساحة معينة في حقل التفاح. وفي حالة عالم الحيوان وحدة التجربة ربما تكون العجول ذات وزن ٦٠٠ رطل أو تكون الثور . والمصطلح الإحصائي الجديد هو متوسط المربعات والذي هو عبارة عن مسمى آخر للتباين.

ولعرض نتائج تحليل التباين فإننا عادة ننشئ جدول بعمود نسميه " متوسط المربعات" والذي يحتوي على التباين لمكونات مختلفة للتحليل.

يركز تحليل التباين على التباين. كما هو معلوم فإنه يمكننا كتابة التباين كمجموع للمربعات مقسوماً على درجات الحرية. وفي تحليل التباين يتم تقسيم إجمالي تباين العينة. إذا كان فرض العدم صحيح فإن الأقسام تعطي تقديرات مستقلة لنفس التباين ويتم حساب إحصاءة توزيع F كنسبة لكلا التقديرين وتوزيع المعاينة لتلك النسبة للتباين مستمرة وتعطي شكل توزيع F (الشكل رقم ٩،١) . إشارة لاسم العالم الذي طورها ويدعى Sir.R.A.Fisher. ويتراوح مدى المتغير العشوائيF من الصفر إلى مالا نهاية ؛ نظراً لأن نسبة الأعداد المربعة لا يمكن أن تكون سالبة. يتميز التوزيع بالتواء موجب ولكنه يصبح متماثلاً كلما اقتربت درجات الحرية لكل تباين من القيمة ما لانهاية. لذا فإننا نتوقع أن تكون أكثر التواء للعينات الصغيرة وأقل التواء كلما زاد حجم العينة.



الشكل رقم (٩.١). توزيع F

وعليه فإن درجة الحرية تحدد شكل التوزيع. ويمكن الاطلاع على الجدول رقم (١٠) بالملحق على قائمة بالقيم الحرجة لـ F لمستويين من المعنوية عندما α تساوي ٥٪ و ١٪ . وتشير القيم في الجدول المكتوبة بالطباعة العادية لقيم F عندما تكون α تساوي٥٪ بينما تشير القيم المكتوبة بالبنط العريض إلى قيم F عندما α تساوي ١٪ . وتشير درجات الحرية الرأسية في الجدول إلى درجات الحرية لتباين البسط في نسبة F بينما تشير درجات الحرية الأفقية في الجدول (الصف) إلى درجات الحرية لتباين المقام.

وتقوم المصطلحات النظرية لتحليل التباين على مجموعه من الافتراضات هي :

١- يتم سحب عينات عددها K عشوائيا من عدد K من المجتمعات.

٢- المجتمعات K مجتمعات طبيعة.

٣- تباينات المجتمعات البالغ عددها K متساوي.

وأيضاً فإن متوسطات المجتمعات K يجب أن تكون متساوية لـ F حتى تكون متغير عشوائي في توزيع F. ولكن هذا الشرط عبارة عن الفرض الذي نختبره باستخدام تحليل التباين. وقبل إجراء الاختبار نعتبر أن الفرض صحيح ولكن ربما يتم رفضه بعد

إجراء الاختبار. أما الشروط الثلاثة الأخرى فيجب أن تكون صحيحة قبل وبعد الاختبار. وللتحقق من صحة الفرض الأول يجب أن نقتنع بأنه تم اختبار العينات عشوائياً باستخدام الطرق التي تم مناقشتها سابقاً لاختبار العينات العشوائية. وهذا الشرط من أهم الشروط الواجب تحققها ؛ لأن مخالفة فرض العشوائية يسبب أخطاء جسيمة في اختبار F أما الشرط الثاني فليس شرط حرج. فإذا كانت المجتمعات غير طبيعية فإن القيم الحرجة الصحيحة لـ F تكون أكبر من تلك الموجودة بالجدول. الشرط الثالث المختص بالتساوي أو التجانس لتباين المجتمعات التي يتم اختيار العينات منها ليس حرجاً أيضاً. فإذا كان التباين لا يختلف كثيراً فإن اختبار F لن يتأثر بدرجة كبيرة ويمكن تخفيض أثر ذلك باختيار عينات ذات حجم متساوي . وبصفة عامة فإنه يوجد طرق إحصائية للتعامل مع الحالات التي تكون فيها المجتمعات غير طبيعية أو ذات تباين غير متجانس.

تحليل التباين عبارة عن موضوع شامل في الإحصاء . ولكن في الجزء التالي سوف يتم مناقشة أسلوب تحليل التباين لحالة تصميم ثلاث تجارب فقط.

تحليل التباين في اتجاه واحد One-Way Analysis of Variance

في تحليل التباين باتجاه واحد يتم تطبيق مجموعة مستويات من المعاملات ثم اختبار فرض العدم القائل بأن متوسطات المعاملات (المعالجات) متساوية مقابل الفرض البديل بعدم تساوي تلك المتوسطات. وهذا مكافئ للقول في فرض العدم بأن تباين متوسطات المعاملات يساوي الصفر مقابل الفرض البديل بأن تباين متوسطات المعاملات أكبر من الصفر. ولذلك يمكن صياغته رياضياً :

$$H_o: \mu_1 = \mu_2 = \ldots = \mu_k (or\, \sigma_\mu^2 = 0)$$
$$H_a: \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k (or\, \sigma_\mu^2 > 0)$$

ويتم رفض فرض العدم إذا كانت F المحسوبة أكبر من الجدولية لمستوي المعنوية α. وعند تساوي المتوسطات فإن عينات عددها K يمكن معاملتها على أنها عينة واحدة من مجتمع واحد، ولكن إذا كانت المتوسطات غير متساوية فإنه يجب التعامل معها على أساس أنها عدد k عينة عشوائية مستقلة مختارة من عدد k مجتمعات مختلفة. وقد ذكرنا ضمن الشروط اللازمة لتحليل التباين بأنه يتم إجراء الاختبار كما لو كان فرض العدم صحيح. لذا، بدأنا بالتعامل مع البيانات كما لو أنها مسحوبة من مجتمع كبير ومتجانس. وقد تم أخذ الاختلاف الكلي وتقسيمه إلى جزأين ، أحدهما للمعاملات والآخر لخطأ التجربة. فإذا كان فرض العدم صحيح ومتوسطات المعاملات متساوية فإن مكونات الاختلاف الكلي للمعاملات لن تحتوي على اختلافات؛ نتيجة اختلاف متوسطات المعاملات وخلاف ذلك فإنه سيكون تقدير آخر لخطأ التجربة.

وإذا تم حساب النسبة لكلا التباينين ، المعاملات وخطأ التجربة طبقاً لهذه الشروط ، فإننا سوف نحصل على قيمه لـ F المحسوبة قريبة من الواحد؛ نظراً لأنه تم تقديرها لنفس الكمية. ولكن إذا كانت متوسطات المعاملات غير متساوية فإن مكونات الاختلاف الكلي للمعاملات سوف تشمل كلا الاختلافين نظراً للاختلاف في متوسط المعاملات وكذلك خطأ التجربة وسيكون متضخم. في هذه الحالة فإن النسبة للتباينين (الاختلافين) ، المعاملات وخطأ التجربة، سوف تعطي رقم كبير لقيمة F المحسوبة — أكبر من القيمة الجدوليه — وسوف تؤدي إلى رفض فرض العدم . وسيتم إجراء هذا الاختبار لحالتين : حالة عينات متساوية الحجم وعينات غير متساوية الحجم.

العينات متساوية الحجم Equal Sample Size

كمثال، انظر للبيانات في الجدول رقم (٩.١) الخاصة بطرق شحن البيض والتي تم تقيمها من قبل تجار الجملة للبيض. تم اختيار أربع طرق للشحن بعدد ٥ شحنات لكل طريقة. وتم تسجيل عدد البيض المكسور في كل شحنه لكل طريقه من الطرق كما في الجدول. فهل متوسط البيض المكسور متساوٍ لكل الطرق المستخدمة في الشحن؟ استخدم مستوى معنوية ١٪ لإجراء الاختبار.

الجدول رقم (٩.١). عدد البيض المكسور لعينات طرق الشحن الأربع.

رقم وحدة المعاينة	A	B	C	D
	طريقة الشحن			
١	٤	٩	١	٩
٢	٣	٧	٦	٧
٣	١	٦	٤	٧
٤	٢	٨	٣	٦
٥	٥	٨	٤	٨
الإجمالي	١٥	٣٨	١٨	٣٧
المتوسطات	$\overline{X}_{.1} = 3.0$	$\overline{X}_{.2} = 7.6$	$\overline{X}_{.3} = 3.6$	$\overline{X}_{.4} = 7.4$

$$\overline{X}_{..} = \frac{15+38+18+37}{5\,(4)} = 5.4$$

طريقة تحليل التباين المستخدمة لتحديد هل طريقه الشحن لها تأثير على عدد البيض المكسور تشتمل على متغيرين. طريقه الشحن وهي عبارة عن متغير نوعي (وصفي) وله أربع قيم، قيمه واحده لكل معاملة ، وعدد البيض المكسور وهو متغير

الاستجابة والذي سيكون دائماً متغير كمي ؛ نظراً لأن الاستجابة المتحصل عليها ربما تعتمد على طريقة المعاملة الخاصة التي استخدمت ؛ وعلى الرغم من أن تحليل التباين يهتم مبدئياً بالعوامل الوصفية فإنه ربما يستخدم أيضاً مع متغيرات كمية إذا تم تحويلها باستخدام بعض الطرق لإنشاء فئات كمية بدلاً من متغيرات مستمرة.

ولمناقشة البيانات في الجدول رقم (٩,١)، فإننا يجب أن نقوم بتعريف بعض الرموز أولاً. يتم تعريف كل رقم لعينة البيض المكسور بالرمز X_{ij} حيث تشير i إلى الصف أو رقم المشاهدة في العينة و j تشير إلى العمود. فمثلاً الرمز X_{21} يساوي ثلاث بيضات مكسورة وهي عدد البيضات المكسورة في الشحنة الثانية التي تم إرسالها باستخدام الطريقة الأولى A . ويمكن حساب متوسط العينة لعدد jth معاملة باستخدام $\overline{X}_{.j} = \Sigma X_{ij}/n$ (j-x) بحيث يتم الجمع لقيم i حيث تمت الاستعاضة بالنقطة بدلاً عن i في $\overline{X}$. وفي المثال الحالي فإن j تساوي ١ ، ٢ ، ٣ أو ٤ بناء على أي المعاملات ندرس. فمثلاً المعاملات الأولى نحسب متوسط العينة بأخذ المجموع الكلي للقيم في العمود A ثم نقسمها على حجم العينة $n = 5$.

$$\overline{X}_1 = \frac{4+3+1+2+5}{5} = 3.0 \quad \text{بيضه مكسورة}$$

وقد تم حساب متوسطات العينات المتبقية بنفس الطريقة وكانت $\overline{X}_{.2} = 7\cdot 6$ ، $\overline{X}_{.3} = 3\cdot 6$ ، $\overline{X}_{.4} = 7\cdot 4$.

والمتوسط العام $\overline{X}..$ (X شرطه نقطه نقطه) عبارة عن المتوسط لجميع المشاهدات العشرين نظراً لأنه إذا كان فرض العدم صحيح فإن كامل البيانات يمكن جمعها. وعليه يمكن حساب $\overline{X}..$ باستخدام المعادلة رقم (9.1):

$$\overline{X}_{..} = \Sigma\Sigma X_{ij}/nk \qquad (9.1)$$

حيث تم الجمع لكلا i ، j وذلك يتطلب استخدام علامتين لصيغة الجمع. يتم الجمع أولاً للصفوف للحصول على العمود الإجمالي ثم نجمع القيم في العمود المتحصل عليه للحصول على المجموع الكلي. ثم نقسم على حاصل ضرب عدد الصفوف n في عدد الأعمدة K وبالتالي تم استخدام جميع مشاهدات العينة في المقام. وبتطبيق ذلك على المثال أعلاه فإن المتوسط الكلي في الجدول رقم (٩،١) يساوي $\bar{X}_{..} = 5 \cdot 4$ بيضه مكسورة.

ونظراً لاستخدامنا مجموعة مجتمعات، فإننا نستخدم بيانات العينة لقياس ثلاثة مصادر للاختلاف في اختلافات المعاملات أو مجموع مربعات المعاملات، والتي تقيس كيفية اختلاف نتائج العينة بين المعاملات المختلفة، اختلافات الخطأ أو مجموع مربعات الخطأ، والتي تقيس بشكل جماعي كيفية اختلاف المشاهدات داخل العينات، والاختلاف الكلي أو مجموع المربعات، والتي تقيس إجمالي الاختلاف لمشاهدات العينة لكامل التجربة بغض النظر عن عدد المجتمعات.

مجموع المربعات للمعاملة Treatment Sum of Squares

لتلخيص الاختلاف بين نتائج العينة المختلفة، يتم استخدام مجموع المربعات للمعاملة والذي يعرف بالرمز SST كما في المعادلة رقم(9.2) :

$$\text{SST} = n\sum(\bar{X}_{.j} - \bar{X}_{..})^2 \qquad (9.2)$$

لحساب SST نجمع مربعات الفرق بين متوسطات المعاملات والمتوسط الكلي ثم نضربها بعدد المشاهدات للعينة n. وعليه فإنه للبيانات في الجدول رقم (٩،١) لدينا :

$$SST = 5[(3.0-5.4)^2+(7.6-5.4)^2+(3.6-5.4)^2+(7.4-5.4)^2]$$

$$SST = 5(5.76+4.84+3.24+4.00) = 89.2$$

ثم ضرب مربع الفروقات عن المتوسط بالرقم٥ $(n = 5)$ عدد المشاهدات للعينة في المعاملة ، وحيث إن عدد المعاملات $K = 4$ ، يكون العدد الإجمالي للمشاهدات ٢٠ مشاهده $(nk = 20)$. ونظراً لأنه تم الحصول على SST كفرق لمتوسطات العينة فإنه يسمى غالبا الاختلاف المفسر. حيث إنه يفسر الفرق في متوسطات العينة الناتج من الفرق الحقيقي في مجتمعات المعاملة بدلاً من الصدفة.

مجموع مربعات الخطأ Error Sum of Squares

يتم حساب الاختلاف داخل العينات ،أو مجموع مربعات الخطأ ، بتجميع مربع الاختلافات بين مشاهدات العينة ومتوسطها. ولذلك يمكن تعريف مجموع مربعات الخطأ SST ، بالمعادلة رقم (9.3) :

$$SSE = \sum\sum(X_{ij} - \overline{X}_{.j})^2 \qquad (9.3)$$

وبحساب هذه القيمة لطرق شحن البيض في المثال السابق الموضح (الجدول رقم ٩،٢) نجد أنها تساوي SSE = 33.6 وتم استخدام نفس التنظيم المستخدم سابقاً لحساب هذه القيمة حيث تم جمع مربع الاختلافات لكل معاملة ثم جمعها للحصول على SSE. وكذلك تم استخدام المشاهدات لكل عينة في حساب مجموع مربعات الخطأ كما فعلنا سابقاً في حساب مجموع مربعات المعاملة.

الجدول رقم (٩.٢). حسابات مجموع مربعات الخطأ لمثال طرق شحن البيض.

القيمة	$(X_{i1} - \overline{X}_{.1})^2$	$(X_{i2} - \overline{X}_{.2})^2$	$(X_{i3} - \overline{X}_{.3})^2$	$(X_{i4} - \overline{X}_{.4})^2$
١	(٤ - ٣)^٢= ١,٠٠	(٩ - ٧,٦)^٢= ١,٩٦	(١ - ٣,٦)^٢= ٦,٧٦	(٩ - ٧,٤)^٢= ٢,٥٦
٢	(٣ - ٣)^٢= ٠,٠٠	(٧ - ٧,٦)^٢= ٠,٣٦	(٦ - ٣,٦)^٢= ٥,٧٦	(٧ - ٧,٤)^٢= ٠,١٦
٣	(١ - ٣)^٢= ٤,٠٠	(٦ - ٧,٦)^٢= ٢,٥٦	(٤ - ٣,٦)^٢= ٠,١٦	(٧ - ٧,٤)^٢= ٠,١٦
٤	(٢ - ٣)^٢= ١,٠٠	(٨ - ٧,٦)^٢= ٠,١٦	(٣ - ٣,٦)^٢= ٠,٣٦	(٦ - ٧,٤)^٢= ١,٩٦
٥	(٥ - ٣)٢^٢= ٤,٠٠	(٨ - ٧,٦)^٢= ٠,١٦	(٤ - ٣,٦)^٢= ٠,١٦	(٨ - ٧,٤)^٢= ٠,٣٦
الإجمالي	١٠,٠٠	٥,٢٠	١٣,٢٠	٥,٢٠

$$SSE = \sum\sum(X_{ij} - \overline{X}_{.j} = 10.00 + 5.20 + 13.20 + 5.20 = 33.6$$

مجموع المربعات الكلي Total Sum of Squares

يتم الحصول على مجموع المربعات الكلي كمجموع لمربعات الفرق بين المشاهدات والمتوسط الكلي كعينه واحده بحيث يتم تجاهل المجموعات في هذه الحالة (معادلة رقم 9.4).

$$Total\ \mathrm{SS} = \sum\sum(\mathrm{X}_{ij} - \overline{\mathrm{X}}_{..})^2 \tag{9.4}$$

ولمثال طرق شحن البيض السابق تم حساب مجموع المربعات الكلي كالتالي :

$$\text{Total SS} = (4 - 5.4)^2 + (3 - 5.4)^2 + \ldots + (8 - 5.4)^2 = 122.8$$

ويجب ملاحظة أنه في حالة جمع مربعات المعاملة مع مربعات الخطأ ، أننا نحصل على مجموع المربعات الكلي ويمكن التعبير عن ذلك رياضياً بالمعادلة رقم (9.5) :

$$\text{Total SS} = \text{SST} + \text{SSE} \tag{9.5}$$

وبتطبيق ذلك على مثال طرق شحن البيض نحصل على :

$$\text{Total SS} = 89.2 + 33.6 = 122.8$$

اختبار F ، The F Test

بحسابنا لمجموع المربعات يمكننا إجراء اختبار F لنقرر رفض فرض العدم من عدمه. $H_0: \mu_1 = \mu_2 = ... = \mu_K$ واختبار F يعبّر عنه كنسبه بالمعادلة رقم (9.6) :

$$F = \frac{\frac{SST}{k-1}}{\frac{SSE}{n-k}} \qquad (9.6)$$

والتي تتبع لتوزيع F بدرجه حرية k-1 للبسط و $n-k$ للمقام حيث k عبارة عن عدد المعاملات و n عدد المشاهدات لكل عينة . ويمكن إعادة كتابة التعبير عنF ، والذي يساوي النسبة بين مجموع المربعات إلى درجات الحرية ، كنسبة لمتوسط المربعات ؛ نظراً لأن تعريف متوسط المربعات عبارة عن مجموع المربعات مقسوماً على درجات الحرية. ولذا يمكن إيضاح ذلك رياضياً كالتالي :

$$F = \frac{MST}{MSE} \qquad (9.7)$$

وتستخدم قيمة F المحسوبة من البيانات باستخدام المعادلة أعلاه كقاعدة لرفض فرض العدم ، فإذا كانت تلك القيمة أكبر من قيمة F_α الجدولية من الجدول رقم (١٠) بالملحق فإنه يتم رفض فرض العدم ، وماعدا ذلك فلا يتم رفضها. وبصفة عامة يتم عرض تحليل التباين باتجاه واحد في جدول يسمى ANOVA كما هو موضح في الجدول رقم (٩،٣) :

الجدول رقم (٩.٣). تحليل التباين باتجاه واحد ANOVA

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	F
المعاملات	$k - 1$	$SST = n\Sigma(\overline{X}_{.j} - \overline{X}_{..})^2$	$MST = SST/(k-1)$	MST/MSE
الخطأ	$k(n-1)$	$SSE = \Sigma\Sigma(X_{ij} - \overline{X}_{.j})^2$	$MSE = SSE/(k(n-1))$	
الكلي	$kn - 1$	$SS = \Sigma\Sigma(X_{ij} - \overline{X}_{..})^2$		

وبذلك فإن جدول ANOVA لمثال طرق شحن البيض يمكن عرضها في الجدول التالي رقم (٩.٤):

الجدول رقم (٩.٤). جدول ANOVA لمثال طرق شحن البيض.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	F
المعاملة	$4 - 1 = 3$	89.2	$89.2/3 = 29.73$	14.16
الخطأ	$4(5-1) = 16$	33.6	$33.6/16 = 2.1$	
الكلي	$4(5) - 1 = 19$	122.8		

ولقبول أو رفض فرض العدم تتم مقارنه قيمة F المحسوبة والتي تساوي $F = 29\cdot73/2\cdot1 = 14.16$ (جدول رقم ٩.٤) بالقيمة الجدولية أو الحرجة F_α بدرجه حرية ٣ للبسط و ١٦ درجه حرية للمقام ومستوى معنوية ١٪ من الجدول رقم (١٠) بالملحق والتي تساوي $F = F_{0.01} = 5\cdot2922$ ، نجد أن قيمة F المحسوبة أكبر من قيمه F الجدوليه ولذلك نرفض فرض العدم في هذه الحالة والتي يمكن القول بناء على ذلك بأن متوسط البيض المكسور غير متساوي في جميع طرق الشحن المستخدمة.

وبدارسة الجدول رقم (٩.٤) نلاحظ أن مجموع المربعات الكلي Total SS يساوي SSE + SST ومجموع درجات الحرية يساوي درجات حرية المعاملات ودرجات حرية الخطأ، وتتحقق هذه العلاقات دائماً في تحليل التباين وتعطي طريقه سهله للتأكد من الخطأ في حالة الحسابات . أيضاً ، درجات الحرية الموجودة في جدول تحليل التباين ANOVA يمكن استخدامها لإيجاد قيم F الجدوليه من الجدول رقم (١٠) بالملحق.

الصيغ الحسابية Computational Formulas

حيث تم عرض الصيغ الخاصة بالمعاملة والخطأ وكذلك المجموع الكلي في الأجزاء السابقة، ولكن لا يفضل استعمال تلك الصيغ لحساب القيم باستخدام الآلة الحاسبة اليدوية.

والطريق الأنسب لحساب قيمه F هو استخدام الكمبيوتر وبرامج أوراق العمل مثل أكسل أو البرامج الإحصائية المتخصصة مثل برنامج SPSS للكمبيوتر الشخصي. ومع ذلك فإننا نعرض هنا المعادلات الحسابية الممكن استخدامها في حالة الحساب بالآلة الحاسبة. وأفضل طريقة لحساب مجموع المربعات الكلي هو استخدام المعادلتين رقمي (9.8 و 9.9) التالية :

$$\text{Total SS} = \sum\sum X_{ij}^2 - C \tag{9.8}$$

$$C = (\sum\sum X_{ij})^2 / kn \tag{9.9}$$

وكذلك يمكن حساب مجموع المربعات للمعاملات باستخدام الصيغ الحسابية الموضحة في المعادلة رقم (9.10) :

$$SST = \sum T_{.j}^{\;2}/n - C \tag{9.10}$$

حيث T_j^2 عبارة عن مجموع المربعات الكلية(العمود) للمعاملة ويمكن حساب مجموع مربعات الخطأ باستخدام الطرح كما هو موضح في المعادلة رقم (9.11) :

$$SSE = \text{Total}SS - SST. \qquad (9.11)$$

ويمكن استخدام البيانات الخاصة بمثال طرق شحن البيض لعرض كيفية عمل الصيغ الحسابية كالتالي :

$$C=(\sum\sum X_{ij})^2\ /kn=(4+3+\ldots+8)^2/(4)(5)=(108)^2/20=583.2$$
$$\text{Total SS} = \sum\sum X_{ij}^{\ 2}-C=(4^2+3^2+\ldots+8^2)-583.2=706-583.2=122.8$$
$$SST=\sum T_{ij}^{\ 2}/n-C=(15^2+38^2+18^2+37^2)/5-583.2=$$
$$SST\ =(225+1444+324+1369)/5-583.2=672.4-583.2=89.2$$
$$SSE=\text{Total SS}-SST=122.8-89.2=33.6$$

وعند إكمال عملية الحساب يمكن إدراج هذه القيم لمجموع المربعات في الأماكن المناسبة لها في جدول تحليل التباين ANOVA كما في العمود الثالث من الجدول رقم (٩,٤) ثم إكمال التحليل كما سبق. ويجب أن نتذكر أن مجموع المربعات لا يمكن أن يكون سالب. ولذلك، يجب أن نعلم أننا قد أخطاءنا في حساب القيم عند الحصول على قيم سالبة لناتج الطرح.

العينات غير متساوية الحجم Unequal Sample Size

حيث إننا في تحليل التباين في الغالب نرغب في الحصول على عينات متساوية الحجم لكل معاملة ولكن في بعض الأحيان لا نستطيع عمل ذلك فمثلاً ، الحيوانات المستخدمة كوحدة للتجربة في تحليل التباين ربما تمرض وتموت، أو اختلاف الطقس ،

الأمراض والحشرات يمكن أن تسبب ضرر للمساحة الحقلية المقامة عليها التجربة ولذلك نفقد بعض المشاهدات ، وربما لأسباب أخرى فقد لا يكون هناك بيانات كافية لبعض المعاملات تسمح بالحصول على عينات متساوية الحجم. في مثل هذه الحالات يمكن القيام بإجراء تحليل التباين ولكن الصيغ المستخدمة لدرجه الحرية ومجموع المربعات مختلفة نوعاً ما.

مجموع المربعات، درجات الحرية،F

Sum of Squares, Degrees of Freedom, and F

مجموع المربعات للمعاملة (المعالجة) معرف كالتالي بالمعادلة رقم (9.12) :

$$\text{مجموع مربعات المعاملة} = SST = \sum n_j(\overline{X}_{.j} - \overline{X}_{..})^2 \qquad (9.12)$$

في هذه الحالة تم ضرب حجم العينة للمعاملة j_{th} بمربع فرق متوسط المعاملة j_{th} من المتوسط الكلي قبل إيجاد المجموع للمعاملات k .

ويحسب مجموع مربعات الخطأ ومجموع المربعات الكلي ، كما هو موضح في المعادلتين رقمي (9.13 و 9.14) :

$$\text{مجموع مربعات الخطأ} = SSE = \sum\sum(X_{ij} - \overline{X}_{.j})^2 \qquad (9.13)$$

$$\text{مجموع المربعات الكلي} = \text{Total SS} = \sum\sum(X_{ij} - \overline{X}_{..})^2 \qquad (9.14)$$

ويمكن حساب درجة الحرية للمعاملات باستخدام الصيغة الرياضية الموضحة في المعادلة رقم (9.15) :

$$\text{درجة حرية المعاملة} \quad df = k - 1 \qquad (9.15)$$

ولحساب درجات الحرية للخطأ نستخدم المعادلة رقم (9.16) والتي تأخذ في الحسبان الفرق في حجم العينات. أيضاً فإن درجة الحرية للإجمالي يمكن حسابها باستخدام الصيغة الرياضية الموضحة في المعادلة رقم (9.17):

درجة حرية الخطأ $\mathrm{df} = \sum (n_j - 1).$ (9.16)

درجة الحرية الإجمالية $\mathrm{df} = (\sum n_j) - 1.$ (9.17)

ويوضح الجدول رقم (٩،٥) جدول تحليل التباين ANOVA المستخدم لحساب إحصاءة F لتحليل التباين باتجاه واحد لبيانات عينات غير متساوية الحجم.

الجدول رقم (٩.٥). تحليل التباين ANOVA باتجاه واحد لعينات غير متساوية الحجم.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	*F*
المعاملة	$k - 1$	$\mathrm{SST} = \sum n_j (\overline{X}_{.j} - \overline{X}_{..})^2$	$\mathrm{MST} = \mathrm{SST}/(k-1)$	MST/MSE
الخطأ	$\sum (n_j - 1)$	$\mathrm{SSE} = \sum\sum (X_{ij} - \overline{X}_{.j})^2$	$\mathrm{MSE} = \mathrm{SSE}/\sum (n_j - 1)$	
الكلي	$(\sum n_j) - 1$	$\mathrm{SS} = \sum\sum (X_{ij} - \overline{X}_{..})^2$		

والصيغة الحسابية المستخدمة في تحليل التباين باتجاه واحد لعينات غير متساوية الحجم تختلف نوعاً ما. وسيتم عرض تلك الصيغ إضافة لمثال توضيحي لطريقة استخدامها في الجزء التالي.

الصيغ الحسابية Computational Formulas

لعرض ذلك نفترض المثال التالي والمتضمن رغبة مدير الإنتاج في مصنع للآلات الزراعية في مقارنة المنتجات لعدد ثلاث آلات إنتاجية. قام المدير بإجراء تجربة للمخرجات لكل آلة كمعاملة وذلك باستخدام نفس الظروف الإنتاجية وقام بتدوين الإنتاج لكل دقيقه. قام بهذه التجربة لمدة ٥ دقائق للآلة A و ١٠ دقائق للآلة B و ٦ دقائق للآلة C و النتائج التي تم التوصل إليها موضحة بالجدول رقم (٩،٦). المطلوب عند مستوى معنوية ٥٪ اختبر هل يوجد فرق في متوسط الإنتاج للآلات الثلاث. ولذلك يمكن صياغة فرض العدم والفرض البديل لهذا الاختبار على النحو التالي :

$$H_0 : \mu_A = \mu_\beta = \mu_C$$
$$H_a : \mu_A \neq \mu_\beta \neq \mu_C$$

ولإجراء اختبار F لهذه التجربة فإنه لابد من حساب مجموع المربعات ويتم ذلك باستخدام الصيغ الحسابية الموضحة في المعادلات من (9.18-9.21) .

$$\text{Total SS} = \sum\sum X_{ij}^2 - C \tag{9.18}$$

$$C = (\sum\sum X_{ij})^2 / \sum n_j \tag{9.19}$$

$$SST = \sum(T_{.j}^2/n_j) - C \tag{9.20}$$

$$SSE = \text{Total SS} - SST \tag{9.21}$$

وباستخدام البيانات الموضحة في الجدول رقم (٩،٦) نحسب أولاً قيمة C ثم مجموع المربعات الكلي :

$$C = (\sum\sum X_{ij})^2 / \sum n_j = 173^2/21 = 1425.19 .$$

$$\text{Total SS} = \sum\sum X_{ij}^{\ 2} - C = (10^2 + 6^2 + \ldots + 12^2) - 1425.19$$

$$\text{Total SS} = 1557 - 1425.19 = 131.81$$

الجدول رقم (٩.٦). مخرجات الإنتاج للآلات الثلاث.

الآلة C	الآلة B	الآلة A
١١	٦	١٠
٨	٧	٦
١٣	٩	٨
١٠	٤	١٢
١٠	٦	٦
١٢	١٠	
	٥	
	٦	
	٨	
	٦	
$\bar{X}_{.3} = 10.7$	$\bar{X}_{.2} = 6.7$	$\bar{X}_{.1} = 8.4$
$T_{.3} = \sum X_{i3} = 64$	$T_{.2} = \sum X_{i2} = 67$	$T_{.1} = \sum X_{i1} = 42$

وبحساب مجموع المربعات للمعاملة نحصل على :

$$SST = \sum(T_{.j}^{\ 2}/n_j) - C = 42^2/5 + 67^2/10 + 64^2/6 - 1425.19$$

$$SST = 352.8 + 448.9 + 682.67 - 1425.19 = 59.18$$

ثم بعد ذلك نستطيع الحصول على مجموع مربعات الخطأ بالطرح كالتالي :

$$SSE = \text{Total SS} - SST = 131.81 - 59.18 = 72.63$$

وبعد إجراء هذه الحسابات فإننا أوجدنا المعلومات الضرورية لتكوين جدول تحليل التباين ANOVA والموضحة بالجدول رقم (٩،٧).

الجدول رقم (٩.٧). جدول تحليل التباين ANOVA لتجربة مخرجات الإنتاج.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	*F*
المعاملة	$3 - 1 = 2$	59.18	59.18/2=29.59	7.33
الخطأ	$\sum(n_j - 1) = 18$	72.63	72.63/18=4.035	
الكلي	$(\sum n_j) - 1 = 20$	131.81		

وباستخدام الجدول رقم (١٠) بالملحق لإيجاد قيمه F الجدوليه عند مستوى معنوية ٥٪ ودرجه حرية للبسط تساوي ٢ ودرجه حرية للمقام تساوي ١٨ نجد أنها تساوي ٣،٥٥. وحيث إن F المحسوبة تساوي ٧،٣٣ من الجدول رقم (٩،٧) وهي تزيد عن F الجدوليه فإن القرار هو رفض فرض العدم. وبالتالي فإنه في حالة هذا المثال فإن المدير يستنتج بأن متوسط مخرجات الإنتاج لهذه الآلات الثلاث غير متساوي.

تحليل التباين باتجاهين Two-Way Analysis of Variance

في حاله تصميم التجارب التي نستخدم فيها مجموعتين من المعاملات أو المعالجات آنياً ربما نختار تجربة قطاع عشوائي بحيث تكون أحد المجموعات ذات أهمية مركزية والمجموعات الأخرى تكون مصدر للاختلاف في البيانات ولذلك يمكن إزالتها

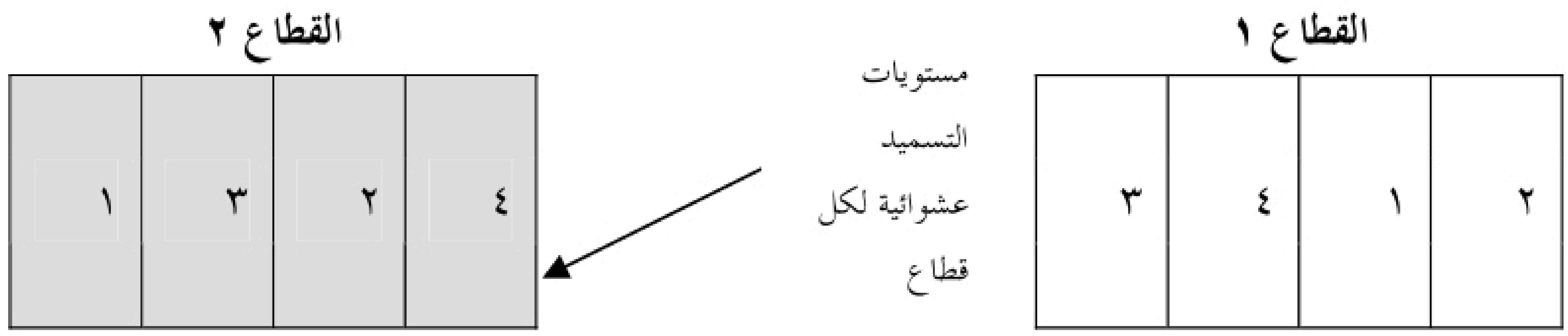
من خطأ التجربة، وربما نقوم بإجراء تجربه تامة التعشية عندما يكون الاهتمام بالاختلاف في كلا العوامل . في هذه الحالة، يكون لدينا مجموعتين من فرض العدم والفرض البديل ونجرى اختبارين لـ F لكل معالجه أو عامل.

تصميم القطاع العشوائي Randomized Block Design

تصميم القطاع العشوائي عبارة عن تصميم عام يستخدم لاختبار t المزدوج. في هذا الاختبار يتم تجميع البيانات ثم التعامل مع الفروق للتخلص من مصادر الاختلاف المعروفة في البيانات.

وفي التصميم التام التعشية يتم إنشاء قطاعات تمثل الوقت والمكان أو أدوات التجربة. ولذلك إذا كان الهدف هو مقارنة ثلاث معالجات وتم الاشتباه بوجود اتجاه زمني في متوسط الاستجابة عبر الزمن، يمكننا التخلص من جزء كبير من اختلاف الاتجاه الزمني باستخدام القطاعات. ويتم عمل ذلك بالتطبيق العشوائي للمعالجات على وحدات التجربة في أحد قطاعات التجربة خلال فتره زمنية معينة ثم إعادة التجربة في فترات زمنيه حتى نجمع البيانات الكافية. أيضاً القطاعات ربما تكون قطع أراضي أو أنواع من الحيوانات أو بعض العوامل الأخرى في التحليل التي يعزى لها جزء من الاختلافات في مجموع المربعات الكلي ويمكن التخلص من أخطاء التجربة بحساب مجموع المربعات للقطاع. فمثلاً إذا كان أحد المهندسين الزراعيين يرغب في تحديد أثر ثلاثة مستويات من تطبيقات الأسمدة على إنتاجية القمح ؛ بالإضافة لقطاع المقارنة بدون إضافة تسميد ويمكن اختيار مواقع مكونه من مزرعتين لإجراء التجربة، واختيار مواقع المزرعة كقطاعات لفصل أثر الاختلاف في خصائص التربة على الإنتاجية عن اختلافات الخطأ. ولذا يمكن أن يكون تصميمه للتجربة مثل الخريطة الحقلية الموضحة في

الشكل رقم (٩،٢) ، بحيث تم اختيار مواقع المعالجات عشوائياً في كل موقع. كل قطاع يجب أن يحتوي على جميع المعالجات و إلا فإنه سيكون لدينا تصميم قطاعي غير مكتمل والذي يكون من الصعوبة تحليله (وأبعد من هدفنا). أيضاً في هذا التصميم فإن المعالجة تظهر مرة واحدة في كل قطاع. والجدير بالذكر فإن هناك تصميم قطاعات عشوائية بحيث يمكن إعادة المعالجات أو تكرارها لكل قطاع ولكن ليست مجال الاهتمام في هذا الكتاب.



الشكل رقم (٩.٢). الخريطة الحقلية لأجهزة التسميد.

تحليل التباين لتصميم القطاع العشوائي

The Analysis of variance for a Randomized Block Design

يحتوي تصميم القطاع العشوائي على متغيرين نوعيين (وصفيين) مستقلين هما القطاعات والمعالجات. لذا فإنه يتم تجزئة مجموع المربعات الكلي إلى ثلاثة أجزاء: المعالجات، والقطاعات، وأخطاء التجربة. نفرض أن $\overline{X}_{i.}$ يعبر عن متوسط القطاع، $T_{i.}$ يعبر عن أجمالي القطاع i th . وعليه فإنه لتصميم قطاع عشوائي به عدد n قطاعات وتم إجراء عدد K معالجات تكون لدينا العلاقات الرياضية التالية والموضحة في المعادلات من (9.22 إلى 9.30).

$$\text{Total SS} = \text{SSB} + \text{SST} + \text{SSE} \quad (9.22)$$

$$\text{Total SS} = \sum\sum(X_{ij} - \overline{X}_{..})^2 \text{ or} \quad (9.23)$$

$$\text{Total SS} = \sum\sum X_{ij}^{\ 2} - C \quad (9.24)$$

$$\text{SSB} = k\sum(\overline{X}_{i.} - \overline{X}_{..})^2 \text{ or} \quad (9.25)$$

$$\text{SSB} = \sum T_{i.}^{\ 2}/k - C \quad (9.26)$$

$$\text{SST} = n\sum(\overline{X}_{.j} - \overline{X}_{..})^2 \text{ or} \quad (9.27)$$

$$\text{SST} = \sum T_{.j}^{\ 2}/n - C \quad (9.28)$$

$$\text{SSE} = \text{Total SS} - \text{SSB} - \text{SST} \quad (9.29)$$

$$\text{Where } C = (\sum\sum X_{ij})^2/kn \quad (9.30)$$

الجدول رقم (٩.٨). جدول ANOVA لتصميم القطاع العشوائي.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	*F*
المعاملات	*k - 1*	*SST*	*MST = SST/(k - 1)*	*MST/MSE*
القطاعات	*n - 1*	*SSB*	*MSB = SSB/(n – 1)*	
الخطأ	*(k – 1)(n – 1)*	*SSE*	*MSE = SSE/(k – 1)(n – 1)*	
الكلي	*kn - 1*	*Total SS*		

ويوضح الجدول رقم (٩,٨) عرض لتحليل التباين لتصميم القطاع العشوائي حيث يحتوي العمود الثاني على درجات الحرية المرتبطة بكل مجموع مربعات. وتم حساب متوسط المربعات بقسمة مجموع المربعات لهم على درجات الحرية الخاصة بها. ويلاحظ وجود قيمة واحدة محسوبة لـ إحصاءة F لهذا الاختبار- وهي عبارة عن متوسط مربع المعالجة مقسوماً على متوسط مربع الخطأ – نظراً لأن فرض العدم الذي

تم اختباره هو هل متوسط المعالجات متساوية أم لا. وقد تم استخدام القطاع في هذه الحالة لتقليل خطأ التجربة.

ويمكن إيضاح ذلك من خلال المثال التالي :

قامت مربية نباتات بتطوير صنف جديد من القمح والذي تأمل أن يعطي إنتاجية أعلى في السهول مقارنة بثلاثة أنواع أصناف مشهورة تزرع تحت ظروف المزارع في الأراضي الجافة. قامت بتصميم تجربه باستخدام القطاعات العشوائية لثلاث أنواع من التربة في المنطقة هي الرملية والطينية والطمي. واختارت حقول لكل نوع من التربة وقسمتها إلى أربعة أجزاء ثم قامت بزراعة الأربعة الأنواع من القمح عشوائياً فيها. ويوضح الجدول رقم (٩,٩) الإنتاجية المتحصل عليها لكل ايكر من كل الأجزاء المزروعة، المطلوب إجراء اختبار لمعرفة هل متوسط الإنتاجية لكل صنف من أصناف القمح متساوية باستخدام مستوى معنوية ٥٪.

الجدول رقم (٩.٩). بيانات إنتاجية قمح المناطق الجافة حسب نوع الصنف والتربة.

القطاع	الصنف				الإجمالي
	A	B	C	D	
رملي	٢٠	٢١	٢١	١٨	٨٠
طيني	٢٥	٢٤	٢١	٢٠	٩٠
طمي	٣٠	٢٨	٢٢	٢٠	١٠٠
الإجمالي	٧٥	٧٣	٦٤	٥٨	٢٧٠

يتم أولاً صياغة فرض العدم والفرض البديل كالتالي :

$$H_o : \mu_A = \mu_B = \mu_C = \mu_D$$

$$H_a : \mu_A \neq \mu_B \neq \mu_C \neq \mu_D$$

ثم نحسب مجموع المربعات باستخدام الصيغ الرياضية التالية :

$$C=(\sum\sum X_{ij})^2/kn=(270)^2/(4)(3)=6,075$$

$$\text{Total SS}=\sum\sum X_{ij}{}^2 - C= 20^2+25^2+...+20^2-6,075=141$$

$$SST=\sum T_{.j}{}^2/n-C=(57^2+73^2+64^2+58^2)/3-6,075=63$$

$$SSB=\sum T_{i}{}^2/k-C=(80^2+90^2+100^2)/4-6,075=50$$

$$SSE = \text{Total SS}-SST-SSB=141-63-50=28$$

ومن النتائج آنفاً نستطيع إنشاء جدول تحليل التباين ANOVA والموضحة في الجدول رقم (٩،١٠) التالي :

الجدول رقم (٩.١٠). جدول ANOVA لاختبار أصناف قمح المناطق الجافة لأنواع محددة من التربة.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	F
المعاملات	(4-1)=3	٦٣	٢١	٤،٥
القطاعات	(3-1)=2	٥٠	٢٥	
الخطأ	(4-1)(3-1)=6	٢٨	٤،٦٧	
الكلي	(4)(3)-1=11	١٤١		

من بيانات الجدول آنفاً نلاحظ بأن قيمة F المحسوبة تساوي ٤،٥٠ والتي تمثل حاصل النسبة TMS/EMS ونقارنها بقيمة F الحرجة والتي يمكن الحصول عليها من الجدول رقم (١٠) بالملحق عند مستوى معنوية ٥٪ وثلاث درجات حرية للبسط و٦ درجات حرية للمقام والتي تساوي ٤،٧٦. وحيث إن قيمه F المحسوبة أقل من قيمة F الجدوليه فإننا لا نستطيع رفض فرض العدم. ولذلك يمكننا القول بأن متوسط إنتاجية القمح في الأراضي الجافة لأصناف القمح الأربعة (بما فيها الصنف الجديد) متساوية.

التصميم العشوائي التام The Completely Randomized Design

التصميم العشوائي التام هو النوع الثاني من تحليل التباين باتجاهين حيث يتركز الاهتمام في دراسة المتغيرين الوصفيين – المعالجات والقطاعات. في هذا التصميم يوجد فرضين للعدم أحدهما للمعالجات والآخر للقطاعات والتي يتم اختبارها مقابل الفروض البديلة الملائمة لكل منهما . لذا فإنه يوجد قيمتان لـ F المحسوبة – واحدة لكل اختبار _ ويتم رفض أو عدم رفض فرض العدم بناء على نتائج الاختبارات الفردية. الصيغ لدرجات ومجموع المربعات ومتوسط المربعات هي نفس الصيغ المستخدمة في حالة تجربة القطاع العشوائي التام. ويمكن عرض ذلك من خلال المثال التالي.

يرغب بستاني الزينة في اختبار مدى استجابة صنف معين من نباتات اليوينسينيا لمستويين من هرمونات النمو وكذلك ثلاثة مستويات من تطبيقات السماد . ويرغب أيضاً في إضافة ضابط لكل معالجة عبارة عن نبات لا يتم معاملته بأي من ذلك. لذا يكون لدينا ثلاث معالجات وأربعة قطاعات. وبعد ثلاثة أسابيع من النمو في البيت المحمي قام بقياس ارتفاع النباتات والموضوعة في حوض مقاس ٨ بوصة. وقد توصل إلى النتائج الموضحة في الجدول رقم (٩.١١).

الجدول رقم (٩.١١). ارتفاع نبات اليوينسينيا (بوصة) المعاملة في حوض ٨بوصة بهرمونات النمو والسماد.

	مستوى هرمونات النمو			
معدل التسميد	الضابطة	الجرعة ١	الجرعة ٢	الإجمالي
الضابطة	٢٤	١٦	١٤	٥٤
منخفض	٢٥	١٧	١٤	٥٧
متوسط	٢٧	١٧	١٥	٥٩
مرتفع	٢٨	١٨	١٦	٦٢
الإجمالي	١٠٥	٦٨	٥٩	٢٣٢

المطلوب الاختبار بمستوى معنوية ٥٪ مدى تأثر متوسط ارتفاع نبات اليوينسينيا بهرمون النمو وتطبيقات السماد.

يتم أولاً صياغة فروض العدم والفروض البديلة لهذا الاختبار كالتالي :

$$H_o:\mu_0 = \mu_L = \mu_M = \mu_H \qquad H_o:\mu_0 = \mu_1 = \mu_2$$
$$H_a:\mu_0 \neq \mu_L \neq \mu_M \neq \mu_H \qquad H_a:\mu_0 \neq \mu_1 \neq \mu_2$$

ولإجراء تحليل التباين لهذا المثال فإن الأمر يتطلب حساب مجموع المربعات ويتم حساب معامل التصحيح Cكالتالي :

$$C=(\sum\sum X_{ij})^2/kn=(232)^2/(3)(4)=4,485.33$$
$$\text{Total SS}=\sum\sum X_{ij}^{\ 2}-C=(24^2+26^2+...+16^2)-4,485.33=310.67$$
$$SST=\sum T_{.j}^{\ 2}/n-C=(105^2+68^2+59^2)4-4,485.33=297.17$$
$$SSB=\sum T_{i.}^{\ 2}/k-C=(54^2+57^2+59^2+62^2)/3-4,485.33=11.33$$
$$SSE=\text{Total SS}-SST-SSB=310.67-297.17-11.33=2.17$$

ويمكن التعويض لتلك النتائج في جدول تحليل التباين ANOVA لحساب قيم F كما في الجدول رقم (٩,١٢).

الجدول رقم (٩.١٢). جدول ANOVA لاختبار أصناف قمح المناطق الجافة لأنواع محددة من التربة.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	F
المعاملات	(3-1)=2	٢٩٧,١٧	١٤٨,٥٨	٤١٠,٨
القطاعات	(4-1)=3	١١,٣٣	٣,٧٨	١٠,٤٥
الخطأ	(3-1)(4-1)=6	٢,١٧	٠,٣٦	
الكلي	(4)(3)-1=11	٣١٠,٦٧		

ولـذلك يمكننـا مقارنـه F المحسوبة والـتي تـساوي ٤١٠,٨ لـهرمـون النمـو بـالقيم الحرجة لـ F التي تساوي ٥,١٤ المتحصل عليها عند مستوى معنوية ٥٪ ودرجة حرية ٢ للبسط ودرجة حرية تساوي ٦ للمقام من الجدول رقم (١٠) بـالملحق لاختبـار الفـرض الأول من فروض العدم. وحيث إن قيمه F المحسوبة أكبر بكثير من قيمه F الجدوليه فإنـه يتم رفض فرض العدم. أما الاختبار الثاني نجد أن قيمه F الحرجة الجدوليه عند مستوى معنوية ٥٪ ودرجه حرية تساوي ٣ للبسط ودرجـه حريـة تـساوي ٦ للمقـام هـي ٤,٧٦ وبمقارنتها بقيمـه F المحسوبة والـتي تـساوي ١٠,٤٥ فإننـا نـرفض فـرض العـدم . ويمكـن تفسير هذه النتائج بأن متوسط ارتفاع نباتات البينستينا تختلف بناء على جرعات هرمون النمو مما يعنى أن لها تأثير على الارتفاع وكذلك الحال بالنسبة للمستويات المختلفـة مـن تطبيقات السماد.

تصميم المربع اللاتيني The Latin Square Design

يوجد بعض الحالات التي تتطلب تحليل ثلاث أو أكثر مـن المعالجـات بنـاء علـى طبيعة المشكلة المدروسة. وتحليل التباين المصمم لمثل هذه الحالات يشار له بنطاق واسع على أنه تصميم عاملي. وأبسط هـذه الحـالات هـو المربـع اللاتـيني والـذي يتعامـل مـع ثلاثة عوامل على أساس مشاهدات قليله نسبياً. ويتم إجـراء التجربـة بتنظـيم مـستويات العامل بإعطائهـا الحـرف A ، B ، ... إلخ في شـكل صـفوف أو مـصفوفة بحيـث يظهـر الحرف مرة واحدة فقط في كل صف وعمود. فعلى سبيل المثال، المربع اللاتيني المكون من أربعة مستويات من المتغيرات يظهر بالشكل التالي (يمكن الرجـوع للملحـق جـدول رقم (١١) للإطلاع على تصاميم أخرى).

D	B	A	C
B	A	C	D
A	C	D	B
C	D	B	A

وهناك عاملان للقطاع، أحدهما للصفوف والآخر للأعمدة، وتتم المعالجات عشوائياً خلالها. هذا التصميم ملائم للمختص بالاقتصاد الزراعي لإجراء الدراسات للمستهلكين بحيث تكون الصفوف عبارة عن مواقع محلات الأغذية في مدينة كبيره والأعمدة عبارة عن أيام الأسبوع . ونعلم من بحوث سابقة أن المستهلكين يشترون في أيام مختلفة من الأسبوع وكذلك أنهم يتسوقون من المحلات القريبة من سكنهم. وحيث إن الناس يسكنون في المناطق بناء على مستوى دخلهم فإن موقع المحل مؤشر مناسب لدخل المستهلك. المعالجة ربما تكون مستوى الأسعار لبرتقال تاكسس، والعدد المباع منه عند كل سعر هي البيانات المطلوب جمعها. ومن جهة أخرى فإن باحث التربة ربما يوظف تصميم المربع اللاتيني لتحديد استجابة إنتاجيه فول الصويا لمعالجات الأسمدة ، ويتم تصميم القطاع باستخدام نفاذية التربة ونسبة الميل للحقل.

وتكون الفروض اللازم اختبارها للمعالجات K هي :

$$H_o : \mu_A = \mu_B = \ldots = \mu_{\mathrm{k}}$$

لعدد k معالجة و :

$$H_a : \mu_A \neq \mu_B \neq \ldots \neq \mu_{\mathrm{k}}$$

ويحدد عدد المعالجات حجم المربع اللاتيني. ويتم عرض البيانات بـ $X_{ij(K)}$ و r^2 للمربع اللاتيني بحيث r صفوف و r أعمدة. الرمز $X_{ij(K)}$ يحتوي على الرمز i الذي يمثل الصف والرمز j الذي يمثل العمود والرمز K الذي يمثل المعالجة.

تحليل التباين للمربع اللاتيني

The analysis of variance for a Latin Square

توضح المعادلات من (9.31-9.36) الصيغ الحسابية المستخدمة في المربع اللاتيني لحساب مجموع المربعات الكلي. حيث تشمل مجموع المربعات الكلي Total SS وصيغه المعالجات (SST) والصفوف (SSR) والأعمدة (SSC) والخطأ (SSE). ويوضح الجدول (٩،١٣) جدول تحليل التباين ANOVA :

$$C=(\sum\sum X_{ij(k)})^2/r^2 \quad (9.31)$$

$$\text{Total SS}=\sum\sum X_{ij(k)}{}^2-C \quad (9.32)$$

$$SST=\sum T_k{}^2/r-C \quad (9.33)$$

$$SSR=\sum T_{i.}{}^2/r-C \quad (9.34)$$

$$SSC=\sum T_{.j}{}^2/r-C \quad (9.35)$$

$$SSE=\text{Total SS}-SST-SSR-SSC \quad (9.36)$$

ويتم إيجاد قيمة F المحسوبة من الجدول رقم (٩،١٣) بينما نوجد القيمة الحرجة لـ F_α وذلك بدرجات حرية (r - 1) للبسط ودرجات حرية (r - 2) (r - 1) للمقام من الجدول رقم (١٠) بالملحق. ويتم رفض فرض العدم عندما تكون قيمة F المحسوبة أكبر من قيمة F الحرجة (الجدولية).

الجدول رقم (٩.١٣). تحليل التباين للمربع اللاتيني.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	*F*
المعاملات	*r - 1*	*SST*	*MST = SST/(r – 1)*	*MST/MSE*
الصفوف	*r – 1*	*SSR*	*MSR = SSR/(r – 1)*	
الأعمدة	*r – 1*	*SSC*	*MSC = SSC/(r – 1)*	
الأخطاء	*(r – 1)(r – 2)*	*SSE*	*MSE = SSE/(r – 1)(r – 2)*	
الكلي	r^2 *- 1*	*total SS*		

الجدول رقم (٩.١٤). تصميم المربع اللاتيني لتجربة تغذية الأبقار.

الصف	العمود		
	I(0%)	II (25%)	III(50+%)
(تبيعة) 1	B 150	A 28	C 256
(ثور) 2	C 288	B 148	A 24
(تبيع) 3	A 60	C 288	B 156

الجدول رقم (٩.١٥). مجاميع الصف والعمود والمعاملات لتجربة تغذية الأبقار.

الصف	العمود			مجموع الصف	مجموع المعاملات
	I(0%)	II (25%)	III(50+%)		
(تبيعة) 1	B 150	A 28	C 256	434	A 112
(ثور) 2	C 288	B 148	A 24	460	B 454
(تبيع) 3	A 60	C 288	B 156	504	C 832
الكلي	498	464	436	1,398	

وكمثال على ذلك نفترض أن تجربة لتغذية الأبقار باستخدام ثلاث معالجات أجريت على حظيرة حيوانات ذات الوزن ٥٥٠ رطلاً بصفوف من الأجناس (تبيعه ، ثور ، تبيع) والأعمدة تساوي نسبة تأثير البراهاما (صفر ، ٢٥ ، أكثر من ٥٠). وتتكون المعالجة من علائق أو وجبات هي ١,٥ (A) ، ٢,٥ (B) ، ٣,٥ (C) في المائة لوزن الجسم والمتغير المقاس هو إجمالي الوزن المكتسب بالرطل خلال فتره التجربة البالغة ١٢٠ يوماً. ويوضح الجدول رقم (٩,١٤) البيانات المتحصل عليها من التجربة. المطلوب اختبار بدرجه معنوية ٥٪ مدى تساوي متوسط الوزن المكتسب للحيوانات للثلاث الوجبات المستخدمة. يمكن صياغة الفروض لهذا المثال على النحو التالي :

$$H_o : \mu_A = \mu_B = \mu_C$$
$$H_a : \mu_A \neq \mu_B \neq \mu_C$$

ولإجراء تحليل التباين فإنه يجب حساب إجمالي الصفوف والأعمدة والمعالجات. ويتم الحصول على إجمالي المعالجات بجمع القيم الثلاث لـ A في الجدول رقم (٩،١٤) للحصول على T_A وكذلك الحال بالنسبة لباقي المعالجات. ويمكن الاطلاع على العمود الأخير في الجدول رقم (٩،١٥) لإجمالي المعالجات.

ولإجراء تحليل التباين فإننا نحسب أولاً معامل التصحيح C وكذلك مجموع المربعات الكلي باستخدام الخطوات التالية:

$$C = (\sum\sum X_{ij(k)})^2 / r^2 = 1{,}398^2 / 3^2 = 217{,}156$$

مجموع المربعات الكلي

$$Total\ ss = \sum\sum X_{ij(k)}{}^2 - c = 150^2 + 288^2 + \ldots + 156^2 - 217{,}156$$
$$Total\ ss = 87{,}968$$

و

$$SST = \sum T_K^2 / r - C = (112^2 + 454^2 + 832^2)/3 - 217{,}156 = 86{,}472$$

بينما

$$SSR = \sum T_{i.}^2 / r - C = (434^2 + 460^2 + 504^2)/3 - 217{,}156 = 834.66$$

و

$$SSC = \sum T_{.j}^2 / r - C = (498^2 + 464^2 + 436^2)/3 - 217{,}156 = 642.66.$$
$$SSE = Total\ SS - SST - SSR - SSC$$

وعليه

$$SSE = 87{,}968 - 86{,}472 - 834.66 - 642.66 = 18.68$$

ويتم استخدام القيم المتحصل عليها لإنشاء جدول تحليل التباين ANOVA والتي يمكن التعبير عنها بالجدول رقم (٩،١٦) التالي :

الجدول رقم (٩.١٦). جدول ANOVA لتجربة تغذية قطيع الأبقار لمدة ١٢٠يوم.

مصدر الاختلاف	درجة الحرية	مجموع المربعات	متوسط المربعات	F
المعاملات	3-1=2	٨٦.٤٧٢.٠٠	٤٣.٢٣٦=٢/٨٦.٧٧٢	٤.٦٢٩
الصفوف	3-1=2	٨٣٤.٦٦	٤١٧.٣٣=٢/٨٣٤.٦٦	
الأعمدة	3-1=2	٦٤٢.٦٦	٣٢١.٣٣=٢/٦٤٢.٦٦	
الخطأ	(3-1)(3-2)=2	١٨.٦٨	٩.٣٤=٢/١٨.٦٨	
الإجمالي	3^2-1=8	٨٧.٩٦٨.٠٠		

ومن النتائج آنفاً نلاحظ أن قيمة F المحسوبة هي ٤٦٢٩ والتي قيمتها أكبر بكثير من قيمة F الحرجة (الجدوليه) والتي تم الحصول عليها من الجدول رقم (١٠) بالملحق عند درجه حرية ٢ للبسط ودرجه حرية ٢ للمقام ومستوى معنوية ٥ ٪ . وعليه فإننا نرفض فرض العدم. ويمكن تفسير ذلك بالقول إن متوسط الوزن المكتسب يومياً غير متساوي للعلائق الثلاث المستخدمة.

ملحق الفصل التاسع Appendix to Chapter 9

تحليل التباين باستخدام برنامج اكسل Analysis of Variance with Excel

ANOVA باتجاه واحد One – Way ANOVA

لاستخدم برنامج اكسل لإجراء تحليل التباين باتجاه واحد ندخل بيانات المعالجة في ورقة العمل بحيث تكون بيانات كل معالجه عمود ثم نختار الأمر تحليل البيانات من قائمة الأدوات . وحيث إنه تم تعريف المعالجات في الصف الأول فإنه يمكن تفسير مدى البيانات في الصفوف الأول في الجدول الذي ظهر لنا . نعلّم الصندوق الذي أمام العلامات ونتأكد أن قيمه ألفا (α) هي ٠.٠٥ ثم نكتب مدى المخرجات في الصندوق الخاص بها ثم نختار

الأمر التنفيذ OK. نحصل بعد ذلك على مخرجات مكونه من جزأين. ويوضح الجدول رقم (٩,١٧) تلك المخرجات حيث يتكوّن من جزأين، الأرقام في الجزء الأعلى تمثل خلاصة البيانات بينما الجزء السفلي يحتوى على تحليل التباين. ويعرّف البرنامج المعالجات بين المجموعات بينما يعرّف الخطأ داخل المجموعات ولكن قيم النتائج المتحصل عليها هي نفسها التي تم إيجادها سابقاً. وقيمة F الجدوليه موضحة في آخر عمود.

الجدول رقم (٩.١٧). مخرجات ورقة العمل لبرنامج اكسل لتحليل التباين باتجاه واحد.

الخلاصة				
المجموعات	**العدد**	**المجموع**	**المتوسط**	**التباين**
A	٥	١٥	٣	٢,٥
B	٥	٣٨	٧,٦	١,٣
C	٥	١٨	٣,٦	٣,٣
D	٥	٣٧	٧,٤	١,٣

ANOVA						
مصدر الاختلاف	***SS***	***df***	***MS***	***F***	***P-VALUE***	***F CRIT***
بين المجموعات	٨٩,٢	٣	٢٩,٧٣٣٣٣	١٤,١٥٨٧	E-059,12	٣,٢٣٨٨٦٧
داخل المجموعات	٣٣,٦	١٦	٢,١			
الإجمالي	١٢٢,٨	١٩				

تحليل التباين ANOVA غير المتساوي
Unequal ANOVA

يعمل تحليل التباين باتجاه واحد في برنامج اكسل بطريقة متساوية مع المعالجات ذات حجم العينات المتساوية وغير المتساوية. ولذلك، ندخل البيانات في ورقة العمل ثم نختار طريقة تحليل التباين باتجاه واحد من تحليل البيانات داخل القائمة أدوات وفي الجدول الذي يظهر لنا نكمل البيانات المطلوبة ثم ننفذ الأمر. وللمثال الذي تم حله سابقاً فإن النتائج باستخدام برنامج اكسل موضحة في الجدول رقم (٩,١٨) التالي :

الجدول رقم (٩.١٨). مخرجات ورقة العمل لبرنامج اكسل لتحليل التباين باتجاه واحد لحالة عدم تساوي حجم العينة.

الخلاصة

المجموعات	العدد	المجموع	المتوسط	التباين
الآلة A	٥	٤٢	٨,٤	٦,٨
الآلة B	١٠	٦٧	٦,٧	٣,٣٤٤٤٤٤
الآلة C	٦	٦٤	١٠,٦٦٦٦٧	٣,٠٦٦٦٦٧

ANOVA

مصدر الاختلاف	SS	df	MS	F	P-VALUE	F CRIT
بين المجموعات	٥٩,١٧٦١٩	٢	٢٩,٥٨٨١	٧,٣٣٢٥٢	٠,٠٠٤٦٨٥	٣,٥٥٤٥٦١
داخل المجموعات	٧٢,٦٣٣٣٣	١٨	٤,٠٣٥١٨٥			
الإجمالي	١٣١,٨٠٩٥	٢٠				

تحليل التباين ANOVA للقطاع العشوائي
Randomized Block ANOVA

لاستخدام برنامج اكسل في تصميم القطاع العشوائي، فإننا ندخل البيانات وليست الإجماليات وذلك بعناوين الصفوف والأعمدة في ورقه العمل ثم نختار من قائمة الأدوات الخيار تحليل البيانات ثم نختار منها ANOVA طريقه عاملين بدون إعادة. يظهر لنا جدول ندخل فيها المدى للبيانات بما في ذلك عناوين الصفوف والأعمدة وذلك في الصندوق الأول ، نعلّم على المربع الذي يشير إلى العلامات، نختار القيمة ٠,٠٥ في الصندوق ألفا (α) ثم نكتب مدى النتائج في الصندوق الأخير وبعد ذلك نختار أمر التنفيذ. تظهر لنا النتائج الموضحة في الجدول رقم (٩,١٩) التالي :

الجدول رقم (٩.١٩). مخرجات ورقة العمل لبرنامج اكسل لتحليل التباين للقطاع العشوائي.

الخلاصة	العدد	المجموع	المتوسط	التباين
رمليه	٤	٨٠	٢٠	٢
طينية	٤	٩٠	٢٢,٥	٥,٦٦٦٦٦٦٧
رملية طينية	٤	١٠٠	٢٥	٢٢,٦٦٦٦٧
صنف A	٣	٧٥	٢٥	٢٥
صنف B	٣	٧٣	٢٤,٣٣٣٣	١٢,٣٣٣٣٣
صنف C	٣	٦٤	٢١,٣٣٣٣	٠,٣٣٣٣٣٣
صنف D	٣	٥٨	١٩,٣٣٣٣	١,٣٣٣٣٣٣

ANOVA

مصدر الاختلاف	*SS*	*df*	*MS*	*F*	*P-VALUE*	*F CRIT*
الصفوف	٥٠	٢	٢٥	٥,٣٥٧١٤	٠,٠٤٦٢٥٨	٥,١٤٣٢٤٩
الأعمدة	٦٣	٣	٢١	٤,٥	٠,٠٥٥٨٤٨	٤,٧٥٧٠٥٥
الخطاء	٢٨	٦	٤,٦٦٦٦٦٧			
الإجمالي	١٤١	١١				

ونلاحظ أن جزء الخلاصة للبيانات يحتوي على الإجمالي والمتوسطات والتباينات بحيث تشمل الصفوف أولاً ثم الأعمدة بعد ذلك. أما الجزء الذي يعطي تحليل التباين ANOVA فإنه يظهر بنفس الطريقة. بحيث يعرض الاختلاف بسبب الصفوف أولاً، وحيث إننا نستخدم الصفوف كقطاعات في هذا التحليل ولسنا مهتمين بهذا الاختبار فإننا نتجاهل قيمة F المحسوبة والجدوليه للصفوف. من جهة أخرى فإننا مهتمين بقيم F المحسوبة والجدوليه للأعمدة؛ نظراً لأننا نهدف إلى اختبار الإنتاجية المرتبطة بالأصناف المختلفة.

تحليل التباين ANOVA للتصميم العشوائي التام

Completely Randomized Design ANOVA

هذه الطريقة مماثله لتصميم القطاع العشوائي باستخدام برنامج أكسل ؛ نظراً لأننا نستخدم نفس طريقه تحليل التباين. ولكن في هذه الحالة فإننا مهتمين بكلا اختبارات F . لذا فإننا في مثال اليونستينا تم أولاً إدخال البيانات بما في ذلك عناوين الصفوف والأعمدة في ورقة العمل وليس الإجماليات ثم نختار من قائمة الأدوات الخيار تحليل البيانات ثم نختار منها الخيار ANOVA : طريقه العاملين بدون إعادة. يظهر لنا جدول ندخل فيه مدى البيانات في الصندوق الأول ثم نؤشر على المربع الذي يشير للعلامات (العناوين) ، وندخل القيمة ٠,٠٥ في الصندوق الخاص بـ ألفا (α) نختار مدى النتائج ثم نختار أمر التنفيذ. تظهر لنا النتائج الموضحة في الجدول رقم (٩,٢٠). ونلاحظ أن قيم F المحسوبة هي نفسها المتحصل عليها سابقاً ماعدا التقريب.

الجدول رقم (٩.٢٠). مخرجات ورقة العمل لبرنامج اكسل لتحليل التباين للتصميم العشوائي التام.

الخلاصة	العدد	المجموع	المتوسط	التباين
السماد الضابطة	٣	٥٤	١٨	٢٨
منخفض	٣	٥٧	١٩	٣٩
متوسط	٣	٥٩	١٩,٦٦٦٦٧	٤١,٣٣٣٣٣
مرتفع	٣	٦٢	٢٠,٦٦٦٦٧	٤١,٣٣٣٣٣
الهرمون الضابطة	٤	١٠٥	٢٦,٢٥	٢,٩١٦٦٦٧
الجرعة ١	٤	٦٨	١٧	٠,٦٦٦٦٦٧
الجرعة ٢	٤	٥٩	١٤,٧٥	٠,٩١٦٦٦٧

ANOVA

F CRIT	*P-VALUE*	*F*	*MS*	*df*	*SS*	مصدر الاختلاف
٤,٧٥٧٠٥٥	٠,٠٠٨٤٨٠٢	١٠,٤٦١٥٤	٣,٧٧٧٧٧٨	٣	١١,٣٣٣٣٣	الصفوف
٥,١٤٣٢٤٩	E-07٣,٧٩٢	٤١١,٤٦١٥	١٤٨,٥٨٣٣	٢	٢٩٧,١٦٦٧	الأعمدة
			٠,٣٦١١١١	٦	٢,١٦٦٦٦٧	الخطاء
				١١	٣١٠,٦٦٦٧	الإجمالي

تمارين Exercises

١- تم إجراء التجربة التالية لتحديد أثر أربعه أنواع من التغذية على الوزن المكتسب للعجول. تم اختيار عدد عشرين عجل وتم تقسيمها عشوائياً في أربع مجموعات وتم تغذية كل مجموعة بنوع من تلك العلائق. ويوضح الجدول التالي الوزن المكتسب بالرطل لكل طريقه . المطلوب باستخدام مستوى معنوية ٥٪ اختبر مدى تساوي متوسط الوزن المكتسب لكل عليقة.

عليقه (٤) Echo Feeds	عليقه (٣) Accu-Ration	عليقه (٢) Peptone	عليقه (١) Ful-O-Vigor
١٩٢	٢٠٧	١٦٠	١٣٠
١٨٧	٢٣٦	١٥١	١٤٧
١٩٥	٢١٦	١٤٨	١٣١
١٩١	٢٣٠	١٥٠	١٥٣
١٩٠	٣٣٧	١٥٥	١٤١

٢- تم تقسيم مجموعة من العجول في مجموعات وفقاً لحجم المزرعة، وتم تسجيل الوزن المكتسب بعد الرعي في حقل القمح لمدة ٦٠ يوماً. كما في البيانات التالية :

مزرعة كبيرة	مزرعة متوسطة	مزرعة صغيرة
١٧٥	١٤٤	١٢٣
١٨٤	١٥١	١١٨
١٦٨	١٥٧	١٣٠
	١٤٨	١٢٥
	١٥٢	
	١٤٠	
	١٥٥	

اختبر بمستوى معنوية ٥٪ مدى تساوي الوزن المكتسب وفقا لأحجام المزرعة.

٣- صنّف أحد الاقتصاديين الزراعيين المزارع متوسطه الحجم التي لم تستفيد من برنامج التمويل الزراعي المحدد وفقاً للنوع والمنطقة كالتالي:

	نوع المزرعة		
المنطقة	قطن	حبوب عليقه	حبوب غذائية
الجنوب الشرقي	٣١	٣٨	٢٧
الوسط	٤٦	٤٢	٣٥
الغرب	٣٠	٢٦	٤٨

اختبر ما إذا كان متوسط الدفع المقدم من البرنامج (بالألف دولار) متساوي حسب نوع المزرعة والمنطقة باستخدام مستوى معنوية ٥٪.

٤- رش البستاني أشجار الجوز تامة النمو في مواقع مختلفة في الحقل باستخدام عدة تطبيقات من الأسمدة خلال فتره أسبوعين لتحديد استجابة الإنتاجية. وتم تسجيل الإنتاجية بالرطل لأشجار الجوز كما في الجدول التالي:

مستوى السماد مضافاً له النيتروجين والزنك			
الموقع	عالي	متوسط	منخفض
بالمجري المائي	١٣٨	١٢٠	١٠٨
منتصف الحقل	٩٠	٧٨	٧٠
اختيار عشوائي	١١٥	٨٩	٧٦

اختبر ما إذا كان متوسط الإنتاجية متساوي لمستويات السماد المختلفة بمستوى معنوية ١٪ مستخدماً المواقع كقطاعات.

٥- اختبر مخططات تصميم التجارب التالية ثم حدد أيهما من نوع المربع اللاتيني.

أ)

A	B	C
C	A	B
B	C	A

ب)

A	B	C
C	C	A
B	A	B

ج)

A	B	C
C	A	B
A	C	A

د)

B	C	D	A
A	D	B	C
D	A	C	B
C	B	A	D

هـ)

B	C	D	A
D	A	B	C
A	C	D	B
C	B	A	D

٦- أكمل جدول تحليل التباين التالي :

مصدر الاختلاف	درجة الحرية	مجموع المربعات	متوسط المربعات	F
المعالجات		١٣٥	٣٣,٧٥٠	
القطاعات	٣		٢٣,٣٣٣	
الخطأ				
الإجمالي		٢٥٠		

٧- يرغب أحد الاقتصاديين الزراعيين في اختبار السوق لمنتج غذائي جديد مصنوع من البطاطس. قام بتصميم مربع لاتيني باستخدام ثلاث معاملات ٠ (وضع السعر عند $A=$ ٠,٩٥دولار ، $B=$ ١,٠٥ دولار ، $C=$ ١,١٥ دولار) وذلك خلال ثلاثة أسابيع في ثلاثة محلات ٠ في ثلاثة مواقع داخلية من المدينة)المطلوب أختبر بمستوى معنوية ١٪ ما إذا كان متوسط المبيعات متساوي لمستويات الأسعار الثلاثة.

الأعمدة

الصفوف	(المحل ١) I	(المحل ٢) II	(المحل ٣) III
أ) (الأسبوع ١)	(C) ٥٠	(B) ٧٩	(A) ١٠٥
ب) (الأسبوع ٢)	(B) ٧٤	(A) ١٠٧	(C) ٦٣
ج) (الأسبوع ٣)	(A) ٩٨	(C) ٦١	(B) ٩٣

٨- استخدم برنامج أكسل لحل التمرين ١ ، ٢ .

٩- استخدم برنامج أكسل لحل التمرين ٣ ، ٤ .

الفصل العاشر

تطبيقات مربع كاي

Chi-Square Applications

في الفصل السابق ، تم مناقشة طريقه اختبارات الفروض حول تساوي ثلاثة متوسطات أو أكثر. والآن سيتم التطرق لدراسة الاختبارات الخاصة بنسب ثلاثة مجتمعات أو أكثر. وتوزيع المعاينة المستخدم في هذه الاختبارات هو مربع كاي ، وهو التوزيع الاحتمالي الذي استخدم لاختبار التباين الأحادي في الفصل السابع. والتطبيقات الأخرى الممكن إجراؤها باستخدام مربع كاي تشمل استخدام بيانات العينة لاختبار استقلالية متغيرين من عدمها ، وكذلك توفيق بيانات العينة بالنسبة لتوزيعات معروفة لمعرفة ما إذا كانت ، على سبيل المثال ، البيانات المدروسة مسحوبة من مجتمع يتبع التوزيع الطبيعي أو التوزيع ذو الحدين ، وسيتم التطرق لهذه التطبيقات في الأجزاء التالية.

اختبارات النسب لعينات عددها K

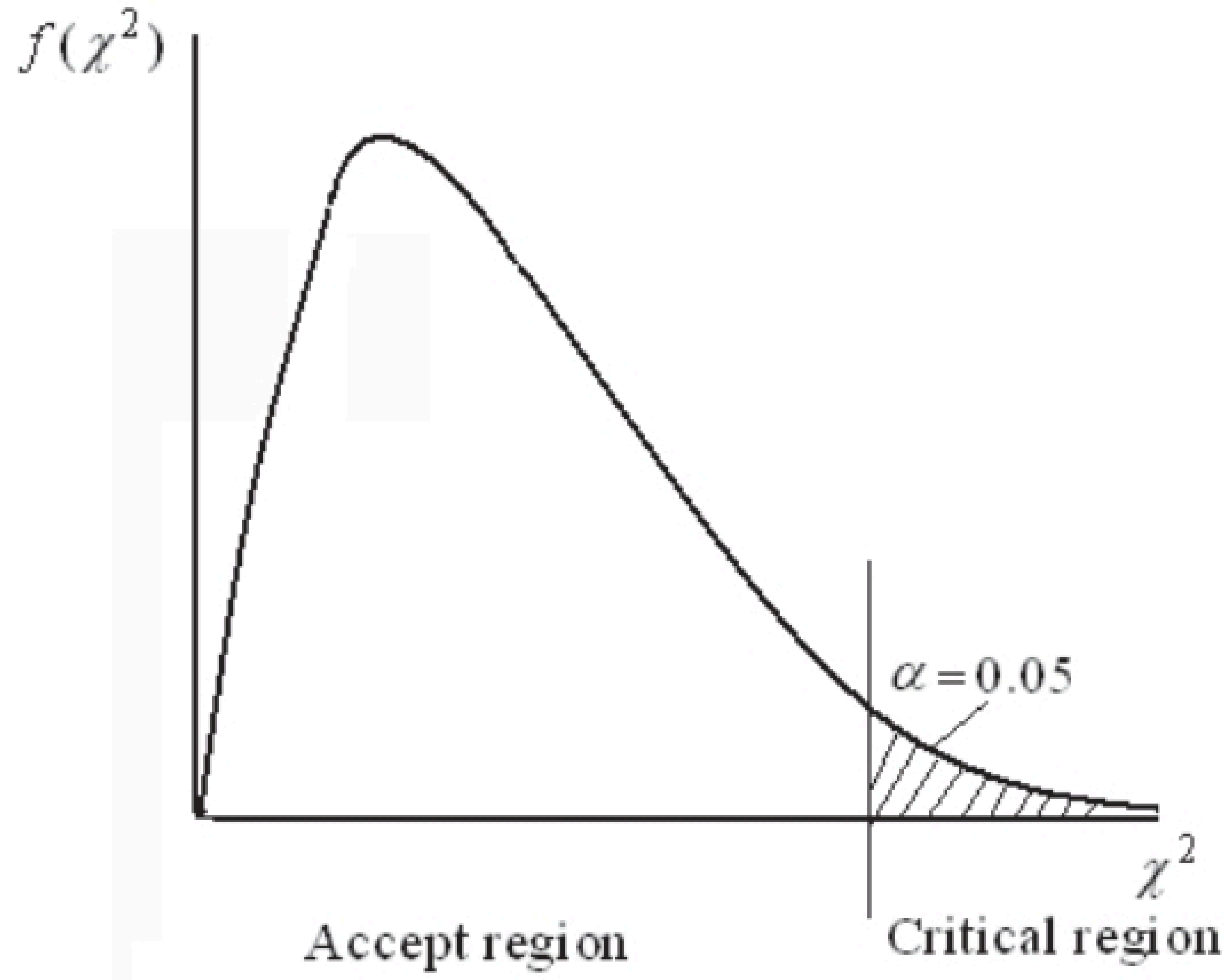
K Sample Tests for Proportions

عند اختيار عينات عددها K ‏(K > 2) من مجتمع ؛ بهدف اختبار فرض العدم حول ما إذا كانت نسب النجاح في هذه العينات متساوية ، فإننا نقوم بإجراء اختبار مربع كاي بطرف واحد هو الطرف الأيمن بدرجه حرية (K - 1). وتشير درجات الحرية

إلى عدد القيود المفروضة على البيانات. ونحن نحتاج إلى عدد المشاهدات في كل عينه وذلك لإيجاد المجموع الكلي للبيانات في كامل المجموعة. وبتحديد درجة الحرية ومستوى المعنوية يمكننا إيجاد القيمة الحرجة لمربع كاي من الجدول رقم (٩) بالملحق. فإذا كانت قيمه مربع كاي المحسوبة من البيانات أكبر من القيمة الحرجة (الجدوليه) فإننا نرفض فرض العدم (الشكل رقم ١٠.١) ويتم حساب قيمة مربع كاي باستخدام الصيغة الرياضية الموضحة في المعادلة رقم (10.1) التالية :

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \qquad (10.1)$$

حيث f_0 عبارة عن تكرار العينة المشاهدة و f_e عبارة عن التكرار المتوقع عندما يكون فرض العدم صحيح.



الشكل رقم (١٠.١). منطقة القبول والرفض لتوزيع مربع كاي عند مستوى معنوية ٥٪.

ويمكن إيضاح ذلك باستخدام المثال التالي. شركة تصنيع الآلات الزراعية لديها ثلاثة مشغلين لنفس الآلة في فترات عمل مختلفة وقام المدير بتسجيل عدد الأجزاء التالفة المنتجة في الفترة الحالية للمشغلين الثلاثة. ويرغب في إجراء اختبار للعينات الثلاث لمعرفة ما إذا كانت نسبة الأجزاء التالفة المنتجة متساوية للمشغلين الثلاثة وسجل المشاهدات في الجدول رقم (١٠،١).

الجدول رقم (١٠.١). اختبار مربع كاي لنسبة الأجزاء التالفة المنتجة بواسطة مشغلي الآلات الثلاثة.

مشغل الآلة	f_0، عدد الأجزاء المشاهدة	f_e، عدد الأجزاء المتوقعة	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
سميث	٢٤	٢٥	- ١	١	٠،٠٤
جون	٣٠	٢٥	٥	٢٥	١،٠٠
قارزا	٢١	٢٥	- ٤	١٦	٠،٦٤
الإجمالي	٧٥	٧٥			١،٦٨

تمت صياغة فرض العدم والفرض البديل للاختبار كالتالي:

$$\mathrm{H_o}: p_1 = p_2 = p_3 = 0.33$$

H_a : عدم تساوي النسب

ويمكن الحصول على عدد الأجزاء التالفة المتوقعة، f_e ، من صيغة مربع كاي المحسوبة الموضحة في الجدول رقم (١٠،١) العمود الثالث بضرب إجمالي عدد التكرارات المشاهدة (٧٥) في النسبة المحددة في فرض العدم (٠،٣٣) . ويوضح الإجمالي في العمود الأخير من الجدول رقم (١٠،١) قيمة مربع كاي المحسوبة

$\chi^2 = 1.68$ ، وباستخدام الجدول رقم (٩) بالملحق ودرجه حرية تساوي ٢ (1-3) ومستوي معنوية ٥٪ نجد أن قيمة مربع كاي الجدوليه تساوي ٥,٩٩١.

وحيث إن قيمة مربع كاي المحسوبة أقل من القيمة الحرجة (الجدولية) فإنه لا يمكننا رفض فرض العدم. ولذلك يمكن القول بأن نسبة الأجزاء التالفة المنتجة بواسطة جميع المشغلين للآلة متساوية.

اختبار الاستقلال في جداول التوافق
Test of Independence in Contingency Tables

في الغالب يتم تلخيص البيانات في جداول إحصائية بحيث يتم عرض عدة مستويات لمتغير واحد في أعمدة الجدول وعرض متغيرات مختلفة أسفل من ذلك (أو عناوين للصفوف). يتركز الاهتمام في اختبار مدى استقلالية تلك المتغيرات. فإذا كانت تلك المتغيرات مستقلة فإن القيم لمتغير الاستجابة (التابع) في متن الجدول لا تعتمد على طريقه التصنيف الذي تم استخدامها في الأعمدة وعناوين الصفوف. ولكن إذا كانت المتغيرات غير مستقلة ، فإن قيم متغير الاستجابة (التابع) لها علاقة بالتصنيف. نستخدم توزيع مربع كاي كقاعدة لهذا الاختبار. وتكون قيمة مربع كاي المحسوبة صغيره عندما تكون متغيرات الصفوف والأعمدة مستقلة و أكبر من القيمة الحرجة لمربع كاي عندما تكون المتغيرات غير مستقلة.

والصيغة المستخدمة لحساب قيمه مربع كاي هي نفسها الصيغة الموضحة في الجزء السابق ولكن يتم حساب درجات الحرية بطريقة مختلفة في هذه الحالة ، يتم ضرب عدد الصفوف مطروحاً منها الواحد بعدد الأعمدة مطروحاً منها الواحد للحصول على درجات الحرية والتي يمكن صياغتها رياضياً في المعادلة رقم (١٠,٢) التالية :

$$v = (r-1)(c-1). \qquad (10.2)$$

ويتم حساب القيمة المتوقعة (f_e) لكل خليه في الجدول اللازمة للتعويض في صيغة مربع كاي المحسوبة والتي تساوي حاصل ضرب إجمالي الصف في إجمالي العمود لهذه الخلية مقسوماً على المجموع الكلي لجميع المشاهدات بالجدول والتي يمكن إيضاحها رياضياً في المعادلة (١٠،٣) التالية :

$$(f_e) = \text{(مجموع الصف)(مجموع العمود)} \; / \; \text{المجموع الكلي} \qquad (10.3)$$

وعدد القيم المتوقعة التي يجب حسابها بهذه الطريقة مساوٍ لعدد درجات الحرية. ويمكن الحصول على باقي القيم المتوقعة بالطرح ؛ نظراً لأن القيم المتوقعة لكل خليه في الجدول يجب إضافتها للحصول على إجمالي الصف والعمود للجدول. ويمكن إيضاح ذلك من خلال المثال التالي.

قام مالك أحد محلات إصلاح الآلات الزراعية المحلية بعملية مسح لأوامر الإصلاح السابقة خلال السنة. وقد لاحظ أن موديلات السنة السابقة للحراثات شملت إصلاحاتها ، الكهرباء ، ومشاكل الوقود ، وأخرى وقد تم تصنيف البيانات إلى مجموعات أو فئات بواسطة شركة الصنع للجرار والمطلوب اختبار هل نوع الإصلاح (العطل) مستقل عن الشركة الصانعة للجرار بمستوى معنوية ٥٪ . والحقيقة إنه في حالة هذا النوع من البيانات ، إذا كان هناك أقل من ٥ مشاهدات لكل خليه ، فإنه لا يمكن استخدام مربع كاي ونحتاج لدمج بعض الخلايا مع بعضها. وعند إكمال البيانات يتم حساب التكرار المتوقع والموضح بين الأقواس في الجدول رقم (١٠،٢). فعلى سبيل

المثال تم حساب التكرار المتوقع للخلية التي تعبر عن الجرار ماركة A وإصلاح الكهرباء كالتالي :

$$(f_e) = (43)(77) / (230) = 14.4$$

الجدول رقم (١٠.٢). عدد الحراثات التي تم إصلاحها حسب نوع الإصلاح وماركة الصنع.

ماركة الصنع	نوع الإصلاح			الإجمالي
	الكهرباء	إمدادات الوقود	أخرى	
A	١٧ (١٤,٤)	١٩ (١٨,٧)	٧ (٩,٩)	٤٣
B	١٤ (١٠,١)	٧ (١٣,٠)	٩ (٦,٩)	٣٠
C	٦ (١٣,٠)	٢١ (١٧,٠)	١٢(٩,٠)	٣٩
D	٣٣ (٣٢,٢)	٤٤ (٤١,٧)	١٩ (٢٢,١)	٩٦
E	٧ (٧,٣)	٩ (٩,٦)	٦ (٥,١)	٢٢
الإجمالي	٧٧	١٠٠	٥٣	٢٣٠

وبعد الحصول على جميع التكرارات المتوقعة، يتم استخدام الصيغة الرياضية الخاصة بحساب قيمة مربع كاي لحساب قيمة الإحصاء ومن ثم مقارنتها بالقيمة الحرجة (الجدولية) لمربع كاي كالتالي :

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(17 - 14.4)^2}{14.4} + \frac{(19 - 18.7)^2}{18.7} + \ldots + \frac{(6 - 5.1)^2}{5.1} = 12.74$$

وتكون درجه الحرية لهذه الحالة، $v = (r - 1)(c - 1) = (5 - 1)(3 - 1) = (4)(2) = 8$.

ومن الجدول (٩) بالملحق و عند مستوى معنوية ٥٪ ودرجة حرية ٨ نجد أن القيمة الحرجة (الجدوليه) لمربع كاي تساوي ١٥,٥٠٧. وحيث أن قيمة مربع كاي المحسوبة $\chi^2 = 12.74$ أقل من القيمة الحرجة فإننا لا نرفض فرض العدم الذي يشير

للاستقلالية. وبالنسبة للمثال فإن ذلك يعني أن ماركه صنع الجرار مستقلة عن نوع الإصلاح، أي أن الإصلاح يتوزع عشوائياً على كل ماركات صنع الجرارات وليس على ماركة واحدة.

في حالة جداول التوافق التي لها صفان وعمودان هناك صيغة رياضية لحساب قيمة مربع كاي، ولكن يفضل استخدام الحاسب الشخصي لإيجادها بدلاً من تعلّم هذه الصياغة.

اختبارات جودة التوفيق Goodness-of–Fit Tests

في معظم المسائل الإحصائية نكون مهتمين بمعرفة مدى مطابقة مجموعة بيانات العينة المشاهدة لتوزيع احتمالي نظري. ويمكن استخدام توزيع مربع كاي كقاعدة لإجراء اختبار بهذه الطريقة. ويتم تطبيق الصيغة الرياضية لمربع كاي المستخدمة في الأجزاء السابقة من هذا الفصل لدراسة ذلك، مع التكرار المتوقع المحسوب كإجمالي لمشاهدات العينة مضروباً بالاحتمال النظري للتوزيع التكراري المطلوب اختباره باستخدام فرض العدم. درجات الحرية لهذا الاختبار تساوي عدد فئات البيانات مطروحاً منها عدد القيود المفروضة على البيانات. فإذا كان من الواجب استخدام البيانات لتقدير القيم للمعالم الضرورية لتعريف التوزيع الاحتمالي النظري المطلوب اختباره، فإننا سوف نفقد درجة حرية لكل معلمة يتم تقديرها إضافة لتلك التي سبق فقدها؛ بسبب حساب إجمالي العينة. فمثلاً، إذا قررنا مقارنة مجموعة من البيانات بالتوزيع الطبيعي فإننا يجب أن نعرف المتوسط والانحراف المعياري قبل إمكانية استخدام صيغة Z لحساب الاحتمالات النظرية للتوزيع الطبيعي. فإذا لم تكن لدينا معلومات خارجية عن تلك القيم فإننا نقوم بتقديرها باستخدام البيانات، لذلك فإننا نفقد درجتي حرية إضافية. لذا فإن لدينا $(K-3)$ درجات حرية لاختبار جودة التوفيق

بدلاً من (K - 1) . التوزيعات الاحتمالية المعتادة التي يتم مقارنه البيانات بها تشمل التوزيع الطبيعي ، ذو الحدين ، بواسون ، والمنتظم. وسيتم التطرق لبعض الحالات والتي تنطبق على أول توزيعين باستخدام أمثلة مناسبة.

في المثال الأول ، شركة حبوب كبرى لديها موظفين يعملون بالساعة عددهم ٨٥٠ وتتوزع أجورهم كما في الجدول رقم (١٠,٣). والمطلوب اختبار ما إذا كانت تلك البيانات تتوزع طبيعياً ، علماً بأن المتوسط والانحراف المعياري التي تم تقديرها لبيانات الأجر بالساعة تساوي ٦,٥ دولار و ٠,٣٥ دولار على التوالي.

الجدول رقم (١٠.٣). بيانات أجر الساعة لموظفي شركة الحبوب.

أجر الساعة (دولار)	عدد الموظفين
٥,٧٥- ٦,٠٠	٦٢
٦,٠٠- ٦,٢٥	١٢٤
٦,٢٥- ٦,٥٠	٢٦٧
٦,٥٠- ٦,٧٥	٢٢٨
٦,٧٥- ٧,٠٠	١٠٦
٧,٠٠- ٧,٢٥	٤٦
٧,٢٥- ٧,٥٠	١٢
٧,٥٠- ٧,٧٥	٥
الإجمالي	٨٥٠

لرسم التوزيع الطبيعي للبيانات نحسب أولاً قيم Z لكل نهاية فئة للتوزيع التكراري في الجدول رقم (١٠،٣)، ثم نوجد الاحتمالات المناظرة لقيم Z المحسوبة ثم نحسب الاحتمال الممكن لشمول الفئة موضع الدراسة على المتغير العشوائي، وعادة يحسب بالطرح. ثم نقوم بكتابه تلك الاحتمالات لكل فئة في عمود في الجدول. وبالحصول على تلك الاحتمالات يتم ضربها بالمجموع الكلي للبيانات (٨٥٠) لحساب التكرارات المتوقعة لهذه الفئات، ثم أخيراً نحسب قيمة مربع كاي باستخدام التكرارات المشاهدة والمتوقعة ثم نقارنها بالقيم الحرجة لمربع كاي.

تتم صياغة الفروض للمثال آنفاً كالتالي:

H_0 : البيانات تتبع التوزيع الطبيعي بمتوسط ٦،٥ دولار وانحراف معياري٠،٣٥

H_a: البيانات لا تتبع التوزيع الطبيعي

والآن سنقوم بحساب القيمة الأولى لـ Z

$$Z = \frac{x - \mu}{\sigma} = \frac{5.75 - 6.50}{0.35} = -2.14, \ P(Z = -2.14) = 0.4838$$
$$Z = \frac{x - \mu}{\sigma} = \frac{6.00 - 6.50}{0.35} = -1.43, \ P(Z = -1.43) = 0.4236$$

وبطرح الاحتمالين من بعضهما نحصل على احتمال أن المتغير العشوائي X يتوزع بين ٥،٧٥ دولار و ٦ دولارات كالتالي:

$$0.4838 - 0.4236 = 0.0602$$

ولكن إذا تم استخدام هذه القيمة فإننا تجاهلنا المساحة التي تقع تحت الطرف الأيسر للتوزيع الطبيعي والتي أقل من القيمة ٥،٧٥ دولار. لذا فإننا نقوم بطرح القيمة ٠،٤٢٣٦ من القيمة ٠،٥٠٠٠ لنحصل على القيمة ٠،٠٧٦٤ والتي تعبّر عن احتمال

الفئة الأولى وبذلك نضعها في الخلية الأولى في العمود ٣ في الجدول رقم (١٠,٤). ولحساب (f_e) لهذه الفئة نضرب إجمالي عدد الموظفين بهذا الاحتمال كالتالي:

$$(f_e) = (850)(0.0764) = 64.9$$

والتي تعطي قيمة الخلية الأولى في العمود رقم ٤ من الجدول رقم (١٠,٤). وسيتم حساب باقي الخلايا في الجدول بنفس الطريقة أعلاه لنحصل على الجدول رقم (١٠,٤) التالي:

الجدول (١٠.٤). الاحتمالات النظرية والتكرارات المتوقعة لبيانات أجر الساعة.

أجر الساعة (دولار)	عدد الموظفين	الاحتمالات الطبيعية	التكرارات المتوقعة
٥,٧٥- ٦,٠٠	٦٢	٠,٠٧٦٤	٦٤,٩
٦,٠٠- ٦,٢٥	١٢٤	٠,١٦٢٤	١٣٨,٠
٦,٢٥- ٦,٥٠	٢٦٧	٠,٢٦١٢	٢٢٢,٠
٦,٥٠- ٦,٧٥	٢٢٨	٠,٢٦١٢	٢٢٢,٠
٦,٧٥- ٧,٠٠	١٠٦	٠,١٦٢٤	١٣٨,٠
٧,٠٠- ٧,٢٥	٤٦	٠,٠٦٠٢	٥١,٢
٧,٢٥- ٧,٥٠	١٢	٠,٠١٤١	١٢,٠
٧,٥٠- ٧,٧٥	٥	٠,٠٠٢١	١,٩
الإجمالي	**٨٥٠**	**١,٠٠٠٠**	**٨٥٠**

والآن يمكن الحصول على قيمة مربع كاي المحسوبة باستخدام الصيغة التالية :

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(62 - 64.9)^2}{64.9} + \frac{(124 - 138)^2}{138} + ... + \frac{(5 - 1.9)^2}{1.9} = 23.84$$

ودرجة الحرية لهذا الاختبار هي $5 = (3 - 8) = (K - 3)$ ، وباستخدام الجدول رقم (٩) بالملحق عند مستوى معنوية ٥٪ ودرجة حرية ٥ نجد أن قيمة مربع كاي الجدوليه تساوي ١١,٠٧. ونظراً لأن قيمة مربع كاي المحسوبة تساوي ٢٣,٨٤ فإننا نرفض فرض العدم ونقول بأن بيانات أجر الساعة لموظفي الشركة الموضحة بالجدول رقم (١٠,٣) لا تتبع التوزيع الطبيعي بمتوسط ٦,٥٠ دولار وانحراف معياري ٠,٣٥دولار. وإذا كان الهدف هو استخدام إجراء إحصائي على هذه البيانات والتي تتطلب أن يكون التوزيع طبيعياً فإننا لسنا قادرين على عمل ذلك.

والمثال الثاني لاختبار جودة التوفيق يتعلق بالتوزيع الاحتمالي ذي الحدين. أحد مربي الأرانب على شكل تجاري قام بجمع البيانات خلال السنة الماضية عند عدد ١٠٠٠ ولادة للأرانب التي كان عدد المواليد لها ٨ لمعرفة عدد الذكور في كل ولادة. وكانت البيانات كما هي موضحة في الجدول رقم (١٠,٥) المطلوب اختبار ما إذا كانت هذه البيانات تتبع التوزيع ذي الحدين بحجم العينة $n = 8$ ، واحتمال $p = 0.5$ عند مستوى معنوية ٥٪ .

أولاً نستخدم الجدول رقم (٢) في الملحق للحصول على الاحتمالات المناظرة لكل قيمه r ونضعها في الجدول رقم (١٠,٥) التالي :

الجدول رقم (١٠.٥). عدد ذكور الأرانب في حظيرة مكونة من ثمانية واحتمالات توزيع ذي الحدين والتكرارات المتوقعة لتحديد جودة توفيق مربع كاي.

عدد الذكورr	العدد المشاهد f_o	احتمالات ذي الحدين	العدد المتوقع f_e
٠	٦	٠,٠٠٣٩	٣,٩
١	٢٧	٠,٠٣١٢	٣١,٢
٢	١١٥	٠,١٠٩٤	١٠٩,٤
٣	٢٢٦	٠,٢١٨٨	٢١٨,٨
٤	٢٧١	٠,٢٧٣٤	٢٧٣,٤
٥	٢١٥	٠,٢١٨٨	٢١٨,٨
٦	٩٨	٠,١٠٩٤	١٠٩,٤
٧	٣٥	٠,٠٣١٢	٣١,٢
٨	٧	٠,٠٠٣٩	٣,٩
الإجمالي	١٠٠٠	١,٠٠٠٠	١٠٠٠

ولحساب قيم (f_e) ، التكرار المتوقع ، نضرب إجمالي الولادات ١٠٠٠ بكل احتمال مناظر ونسجل النتائج في الجدول أعلاه.

ومن النتائج المتحصل عليها في الجدول نستطيع حساب قيمه مربع كاي باستخدام الصيغة الرياضية الخاصة كالتالي:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(6-3.9)^2}{3.9} + \frac{(27-31.2)^2}{31.2} + \ldots + \frac{(7-3.9)^2}{3.9} = 6.4$$

وباستخدام جدول مربع كاي رقم (٩) بالملحق عند درجه حرية ٨ ومستوى معنوية ٥٪ نجد أن قيمة مربع كاي الحرجة (الجدوليه) تساوي ١٥,٥٠٧.وبمقارنة قيمة

مربع كاي المحسوبة بالقيمة الحرجة (الجدوليه) نجد أن القيمة الجدوليه أكبر مما يعني أننا لا نستطيع رفض فرض العدم القائل بأن البيانات تتبع التوزيع ذي الحدين بـ $n = 8$ ، $p = 0.5$. ولذلك فإن عدد الذكور في ولادات الأرانب التي تشتمل كل منها على ٨ صغار لهذا المنتج كما هو متوقع بأن نسبة الذكور فيها يجب أن تكون ٠,٥ .

تمارين Exercises

١- يرغب باحث تسويق في معرفة مدى تفضيل المستهلكين لمذاق أربعة أنواع من عصير البرتقال هي : العصير الطازج ، المجمد ، المصنوع من البودرة والمعلّب. تم اختيار عينة حجمها ١٠٠ مستهلك للعصير وتم إعطاء كل مستهلك أربعة أكواب من العصير تم ترقيمها من A إلى D على التوالي وتم سؤاله أي الأنواع يفضل وتم الحصول على البيانات التالية :

النوع المفضل	عدد المستهلكين
A	٣٣
B	٢٩
C	٢١
D	١٧

اختبر بمستوى معنوية ٥٪ ما إذا كانت نسبة التفضيل لكل نوع متساوية.

٢- نوعان من الزهور المهجنة تم تهجينها بواسطة البستاني وكانت النتيجة المتحصل عليها بأن ألوان التويج للنبات كالتالي : ١٢٢ أرجواني، ٥٠ زرقاء ، ٣٨ حمراء، ١٤ منقطة. هل تتعارض هذه النتائج مع نسبة التكرار النظري التالية ٩ : ٣ : ٣ : ١ ؟ استخدم مستوى معنوية ٠,٠١.

٣- تمت معاملة بذور البطاطس كيميائياً ضد مرض اللفحة كما في الجدول التالي :

المعاملة	البذور الحية	البذور الميتة
تم معالجتها	١٥٢	٤٨
لم تعالج	٣٣	٦٧

هل عدد نباتات البطاطس التي عاشت مستقلة عن المعالجة التي تمت لها ؟ استخدم مستوى معنوية ٥٪.

٤- أحد وكلاء الإرشاد الزراعي في المحافظة تم تكليفه بمهمة جديدة وبناءً على مهامه الجديدة في الوظيفة قام بالاطلاع على بيانات عدد منتجي الثروة الحيوانية في المحافظة حسب فئاتهم طبقاً للعمل اليومي هل يعمل بدوام كامل أم دوام جزئي وربط هذه المعلومات بالسجلات الموجودة لديه في المكتب والخاصة بالتحاقهم ببرامج التعليم الموجهة لمالكي الثروة الحيوانية المقدمة من المركز. فكانت النتائج كالتالي :

نوع المنتج	التحق بالدورة	لم يلتحق
دوام كامل – قطيع كبير	١٧	٣٣
دوام كامل – قطيع متوسط	٢١	١٤
دوام جزئي – قطيع صغير	٣٢	١٣

هل الالتحاق بالدورة مستقل باستخدام مستوي معنوية ١٪.

٥- البيانات التالية توضح سقوط المطر السنوي في موقع معين خلال فترة زمنية ٦٥ سنة. اختبر بمستوي معنويه٥٪ هل سقوط المطر يتوزع طبيعياً بمتوسط ١٨ بوصة وانحراف معياري ٧,٩.

التكرار	معدل سقوط المطر (بوصة)
٩	أقل من ١٠
١٤	١٠- ١٥
١٨	١٥- ٢٠
١١	٢٠- ٢٥
٨	٢٥- ٣٠
٥	٣٠ أو أكثر

٦- اختبر ما أذا كان عدد المزارعين الواصلين لمحلج القطن للتفريغ خلال فترة ١٥ دقيقة يتبع لتوزيع بواسون بمتوسط يساوي ٢ باستخدام مستوى معنوية ٥٪.

عدد فترات الـ ١٥ دقيقة	عدد المزارعين
٢٢	صفر
٥٠	١
١٦	٢
١٠	٣
٧	٤ أو أكثر

٧- يرغب أحد علماء الإنتاج الحيواني في نشر النتائج التي تحصل عليها حول تأثير تغذية خاصة على الاختلاف في حجم المواليد في الأرانب. في دراسته أختبر عشوائياً ٢٥ أنثى من الأرانب وقام بتغذيتها خلال فتره الحمل. كان متوسط وتباين تلك العينة هو $\overline{X} = 8.3$ ، $S^2 = 3.4$.

المطلوب إيجاد فتره الثقة ٩٠٪ لتباين المجتمع لهذه المواليد بناء على هذه التغذية.

٨- في قسم التحكم بالجودة في مصنع الدباغة تم أخذ عينة حجمها ٥ من خط الإنتاج وتم تسجيل عدد المعيب منها. وبعد سحب عينات حجمها ١٠٠٠ تم الحصول على البيانات التالية:

عدد الجلود المعيبة	عدد العينات
صفر	٧٥٢
١	٢١٠
٢	٢٥
٣أو أكثر	١٣

اختبر ما إذا كانت هذه النتائج تتبع التوزيع ذي الحدين.

الفصل الحادي عشر

الارتباط، الانحدار، والسلاسل الزمنية

Correlation, Regression, and Time Series

حتى الآن قمنا باستخدام المعلومات المتاحة في البيانات التي تم جمعها للمتغير العشوائي X لعمل استدلال حول القيم التي نرغب في معرفتها للمتغير. وبذلك تم اختبار المتوسط لـ X وذلك باستخدام المتوسط كمؤشر لمعظم القيم في مجموعة البيانات. الآن يتركز الاهتمام على المعلومات المتاحة في متغير آخر Y ونختبر لنرى ما إذا كان هناك علاقة بين المتغيرين X و Y والتي يمكن أن تعطي معلومات أكثر حول المتغير X بدلاً من المتوسط فقط. وفي حالة وجود زوجين من المتغيرات فإن لدينا نوعين من التحليل ممكنة. أحد تلك الأنواع، عشوائية كلا المتغيرين و يتركز الاهتمام في معرفة درجة العلاقه بينهما. فإذا كانت درجة العلاقة بينهما قويه بحيث إن زيادة قيمه أحدهما تؤدي لتوقع زيادة الآخر ولا يمكننا ملاحظة إلا القيم الخاصة بأحد تلك المتغيرين فإنه يمكن استخدامها للمساعدة في تقدير الكميات الخاصة للمتغير الذي لا نستطيع ملاحظته. أما النوع الآخر من التحليل فإن أحد تلك المتغيرات عشوائي بينما المتغير الآخر ثابت وعليه فإنه يمكننا معرفة القيم المفترضة بدرجه معينه من اليقين. وإذا كان بالإمكان وصف العلاقة بينهما رياضياً فإن لدينا طريقه تمكننا من تقدير القيم

المتوقعة للمتغير العشوائي باستخدام قيم معينة للمتغير الثابت. وتجدر الإشارة إلى أن النوع الأول من التحليل هو الارتباط بينما النوع الثاني من التحليل هو الانحدار وسيتم التطرق لهما في الأجزاء التالية.

تحليل الارتباط Correlation Analysis

ارتباط متغيرين عشوائيين X ، Y يعتمد على ما إذا كانت تتغير في نفس الاتجاه أو في الاتجاه المعاكس أو في اتجاه غير محدد. والحالة الأخيرة تعبّر عن استقلالية المتغيرات وليس ارتباطها . فعلى سبيل المثال أسعار القمح تكون مستقلة عن أسعار جهاز الحاسب المكتبي ؛ نظراً لأنها تتحدّد وفقاً لقوى سوقية مختلفة تماماً عن بعضها. وإذا حصل وأن تغيرت الأسعار معاً خلال فتره قصيرة فيمكن اعتبار أنه حدث عشوائي قصير الأمد. ولكن أسعار القمح وأسعار الشعير غالباً تميل للتغير في نفس الاتجاه ؛ بسبب أنها تتأثر بنفس القوى السوقية وبينها ارتباط موجب. وبالمثل ، فإن سعر وكمية القمح المستهلكة متغيرات عشوائية تتغير في الاتجاه المعاكس وبينها ارتباط سالب. ويمكن التفكير في العديد من الأمثلة المتعلقة بالمتغيرات العشوائية المرتبطة وغير المرتبطة.

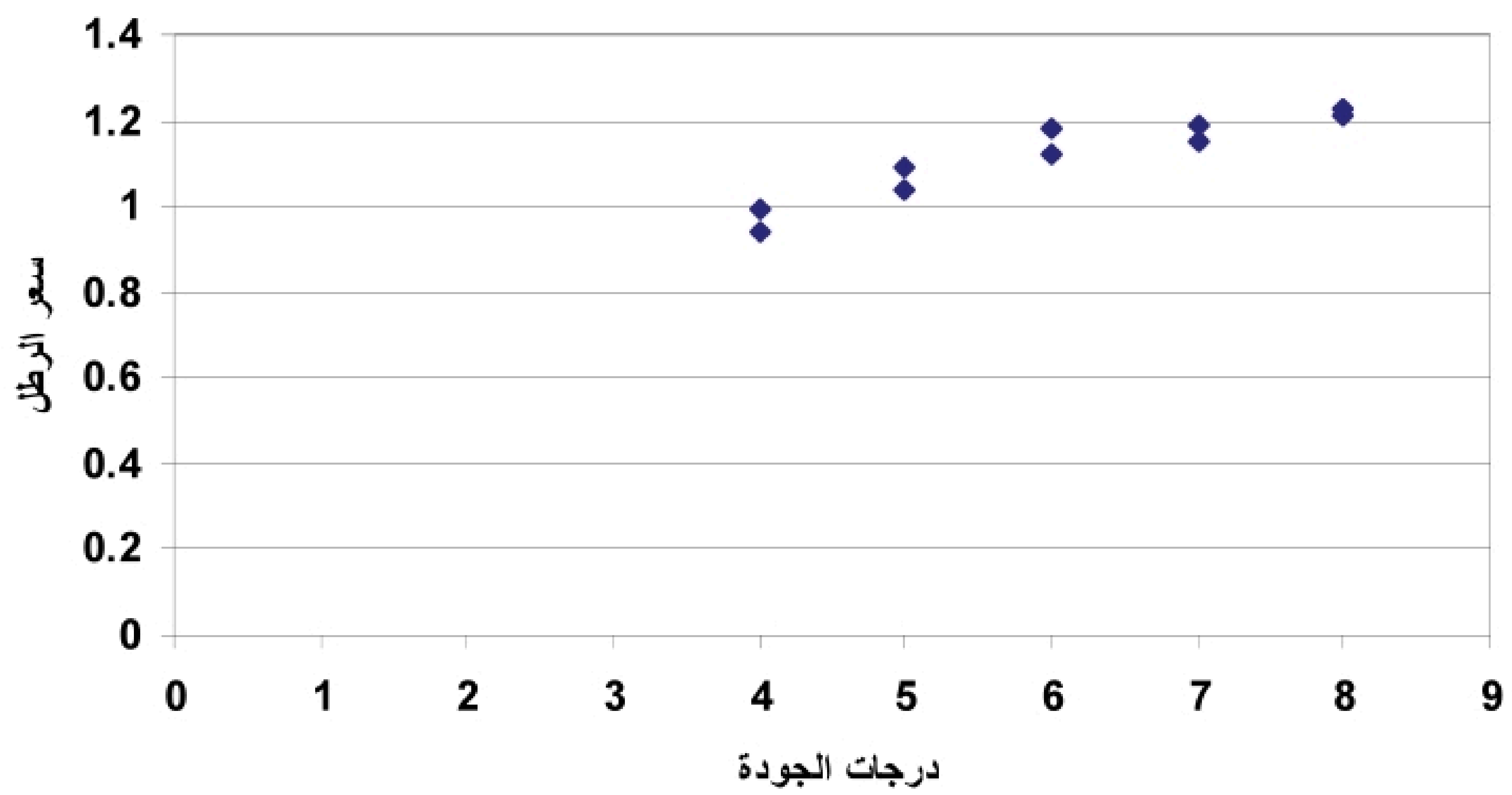
يقيس معامل الارتباط r قوة درجه الارتباط بين متغيرين عشوائيين. وتتراوح قيمته بين ±١ ، أي $-1<r<+1$. وتكون قيمه هذا المعامل للمتغيرات العشوائية المستقلة قريبة من الصفر بينما تكون قيمته موجبة للمتغيرات المرتبطة في نفس الاتجاه وسالبة للمتغيرات التي بينها علاقة عكسية. وكلما كانت القيمة المطلقة لهذا المعامل r كبيره كانت درجة الارتباط بين المتغيرين قويه. ويمكن حساب معامل الارتباط باستخدام أي من الصيغ الرياضية التالية ، المعادلتين رقمي (11.1 و 11.2) :

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}} \quad (11.1)$$

$$r = \frac{\sum XY - n\overline{X}\overline{Y}}{\sqrt{(\sum X^2 - n\overline{X}^2)(\sum Y^2 - n\overline{Y}^2)}} \quad (11.2)$$

ونظراً لأننا نأخذ القيمة الموجبة للجذر التربيعي للمقام دائما لكلا المعادلتين ، فإن r تكون سالبة في حالة كان حاصل طرح البسط سالب.

ويمكن استخدام شكل الانتشار لعرض الارتباط بين المتغيرين العشوائيين X ، Y بيانياً . في شكل الانتشار نرسم كل زوج (X_i, Y_i) بنقطة لكل البيانات ثم نختبر الشكل المتحصل عليه (الشكل رقم ١١.١) . وتقع كل النقاط على خط مستقيم إذا كانت $r = 1$ ويكون الخط موجب الميل إذا كانت r موجه وسالب الميل إذا كانت r



الشكل رقم (١١.١). شكل الانتشار لبيانات شرائح ثور لانجوس.

سالبة. وعندما تكون r أقل من ١ فإن النقاط تقع في نطاق ضيق أو على شكل قطع ناقص والتي تصبح أقرب للخط كلما قربت قيم r من الواحد. وعندما تكون قيمه r مساوية للصفر فإن النقاط تكون على شكل دائرة أو كتلة بدون اتجاه محدد. القيم الصغيرة لـ r تعطي شكل انتشار بنقط أقرب للدائرة منها للخط المستقيم.

ويمكن اختبار معامل الارتباط باستخدام توزيع t . حيث يمكن صياغة فرض العدم والفرض البديل كالتالي :

$$H_0 : \rho = 0$$
$$H_a : \rho < 0 \text{ (or } \rho > 0)$$

حيث ρ (حرف إغريقي ينطق رو) يعبّر عن معامل الارتباط للمجتمع. درجه الحرية للاختبار هي $(n - 2)$ ؛ نظراً لأننا نخسر درجات حرية لتقدير متوسطات المتغيرين X و Y ، و n تعبر عن عدد أزواج البيانات في المجموعة وتحسب قيمة t باستخدام المعادلة رقم (11.3)، وحيث إن $\rho = 0$ في فرض العدم، فإننا عادة لا نضعها في البسط عند حساب قيمة t .

$$t = \frac{r - \rho}{S_r} = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} \qquad (11.3)$$

ويمكن إيضاح ذلك من خلال المثال التالي :

تم تحديد درجات الجودة لنوعيه لحوم البقر المذبوحة بواسطة قسم الزراعة الامريكيه كالتالي :

جيد منخفض (slightly abundant marbling ؛ الدرجة=٨)، اختيار مرتفع

(moderate marbling ؛ الدرجة =٧)، اختيار متوسط(modest marbling ؛ الدرجة = ٦)، اختيار منخفض (small amount of marbling ؛ الدرجة = ٥)، اختيار (slight marbling ؛ الدرجة = ٤)، مستوى عالي (trace of marbling ؛ الدرجة = ٣)، ومستوى متوسط (practically devoid of marbling ؛ الدرجة = ٢).ويرغب المختص في الإنتاج الحيواني في تحديد درجة العلاقة بين درجه نوعية اللحوم المحدّدة من قبل إدارة الزراعة الأمريكية *(X)* وسعر البيع للرطل من تلك القطع *(Y)* وذلك لعينة عشوائية حجمها ١٠ قطع من لحم العجل والبيانات موضحة بالجدول رقم (١١.١). ويمكن

الجدول رقم (١١.١). درجات جودة أنواع شرائح لحوم البقر المذبوحة وأسعارها لشرائح ثور لانجوس.

درجة جودة النوع، X	سعر البيع للرطل بالدولار، Y	XY	X^2	Y^2
٥	١,٠٩	٥,٤٥	٢٥	١,١٨٨
٧	١,١٩	٨,٣٣	٤٩	١,٤١٦
٤	٠,٩٩	٣,٩٦	١٦	٠,٩٨٠
٥	١,٠٤	٥,٢٠	٢٥	١,٠٨٢
٨	١,٢١	٩,٦٨	٦٤	١,٤٦٤
٦	١,١٨	٧,٠٨	٣٦	١,٣٩٢
٤	٠,٩٤	٣,٧٦	١٦	٠,٨٨٤
٧	١,١٥	٨,٠٥	٤٩	١,٣٢٢
٨	١,٢٣	٩,٨٤	٦٤	١,٥١٣
٦	١,١٢	٦,٧٢	٣٦	١,٢٥٤
٦٠	١١,١٤	٦٨,٠٧	٣٨٠	١٢,٤٩٥

حساب قيمة r بالتعويض عن القيم الموجودة في الجدول في واحدة من المعادلات الحسابية وحلها لنجد أن قيمة $r = 0.94$.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}} = \frac{68.07 - \frac{(60)(11.14)}{10}}{\sqrt{\left[380 - \frac{(60)^2}{10}\right]\left[12.495 - \frac{(11.14)^2}{10}\right]}} = 0.94$$

والتي تشير إلى أن درجه الارتباط بين درجه نوعية اللحم وسعر الرطل منها مرتفعة جداً. ويمكن اختبار هذه القيمة لمعرفة ما إذا كانت فعلاً أكبر من الصفر باستخدام اختبار t والفرض المناسب هو :

$$H_0 : \rho = 0$$

$$H_a : \rho > 0$$

ونستخدم اختبار من طرف واحد باتجاه اليمين لهذا المثال ؛ نظراً لأننا نتوقع وجود ارتباط موجب بين المتغيرين العشوائيين ، درجه جودة اللحوم وسعرها. درجة الحرية لهذا المثال هي :

$$v = (n - 2) = (10 - 2) = 8$$

وباختيار ٥٪ مستوى معنوية فإن قيمة t الحرجة (الجدوليه) من جدول رقم (٨) بالملحق هي ١,٨٦ . فإذا كانت قيمة t المحسوبة أكبر من ١,٨٦ فإننا نرفض فرض العدم ، خلاف ذلك لا نرفضه. وبحساب قيمه t باستخدام الصيغة الرياضية نجد أنها تساوي ٧,٧٩ كالتالي :

$$t=\frac{r-\rho}{\sqrt{\frac{1-r^2}{n-2}}}=\frac{0.94-0}{\sqrt{\frac{1-0.94^2}{10-2}}}=7.79$$

وحيث إن قيمة t المحسوبة $t=7.79$ أكبر من قيمة t الجدوليه $t=1.86$ فإننا نرفض الفرض القائل بعدم وجود ارتباط بين المتغيرين - درجه جودة اللحوم وسعرها للرطل. ولذلك فإن المتغيرين تتحرك في نفس الاتجاه؛ كما في الشكل رقم (١١،١).

الانحدار الخطي البسيط Simple Linear Regression

صُمِّم تحليل الانحدار للإشارة إلى العلاقة الرياضية بين المتغير الثابت X والمتغير العشوائي Y في معادلة يكون فيها Y متغير تابع، لذا فإن المعادلة التي لدينا تكون على الشكل $Y=a+bX$. وهذه المعادلة خطية في المعالم a ، b حيث a تشير إلى القاطع بينما b عبارة عن الميل إذا كانت المعادلة توضح خط مستقيم. ولمعرفة شكل الدالة يجب جمع بيانات على شكل أزواج مرتبة (X_i, Y_i) واستخدامها للحصول على تقدير لمعالم الدالة a ، b. فمثلاً الخط $Y = 42 - 2X$ يختلف عن الخط $Y = -100 + 0.7X$ واختلافها واضح كلياً من قيم a ، b. رياضياً ، ربما نستخدم العديد من الطرق للحصول على قيم a ، b باستخدام البيانات الخاصة X ، Y . ولكن طريقة المربعات الصغرى والتي تركز على المتغير العشوائي Y تستخدم لتحليل الانحدار. وبصفة خاصة فإن طريقه المربعات الصغرى تعمل على تدنية مجموع مربعات الانحرافات لكل نقطه من البيانات (X_i, Y_i) من الخط الذي يتم تقديره في اتجاه Y . وحيث إنه من غير المتوقع أن تقع كل نقاط البيانات الأصلية (X_i, Y_i) على الخط المستقيم ، لكل قيم X_i في البيانات، لذا فإنه قد يكون لدينا قيمتين لـ Y. القيمة الأصلية من البيانات Y_i والقيمة المتحصل عليها من المعادلة Y_e

والتي تعتمد على قيم a ، b . والفرق بينهما $(Y_i - Y_e)$ هو الاختلاف محل النقاش. وحيث إن المجموع $\sum(Y_i - Y_e) = 0$ فإنه بذلك غير مفيد لنا، ولكن المجموع $\sum(Y_i - Y_e)^2$ لا يساوي الصفر ويمكن اختياره كمعيار لرسم الخط لمجموعة البيانات. وعليه فإن طريقه المربعات الصغرى تعمل على تدنية $\sum(Y_i - Y_e)^2$. رياضياً ، نحن نعمل على تصغير هذا المجموع بأخذ التفاضل لها بالنسبة لـ a ، b ثم مساواتها بالصفر وحل المعادلة . ولذلك نحصل على ما يسمى بالمعادلات الطبيعية المعادلة رقم (11.4) .

$$\begin{aligned} \sum Y_i &= na + b\sum X_i \\ \sum X_i Y_i &= a\sum X_i + b\sum X_i^2 \end{aligned} \tag{11.4}$$

ويمكن حل المعادلة الأولى بالنسبة لـ a للحصول على الصيغة التعريفية والصيغة الحسابية (معادلتي رقم 11.5 و 11.6).

$$a = \overline{Y} - b\overline{X} \tag{11.5}$$

$$a = 1/n(\sum Y_i - b\sum X_i) \tag{11.6}$$

وبالتعويض عن الطرف الأيمن من المعادلة الأخيرة في المعادلة الطبيعة الثانية والمعالجة الرياضية لها نحصل على تعبيرين رياضيين لـ b (المعادلة رقم 11.7) :

$$b = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} = \frac{\sum X_i Y_i - \dfrac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^{\,2} - \dfrac{(\sum X_i)^2}{n}} \tag{11.7}$$

وتجدر الإشارة إلى أن المعادلة المشتملة على متوسط X ومتوسط Y أسهل في الاستخدام بالآلة الحاسبة اليدوية إذا كانت كلا المتوسطين أرقام صحيحة ، وما

عدا ذلك فإن التعبير الثاني لـ b يكون أسهل استخدام. وتحليل الانحدار يمكن عمله بكفاءة باستخدام البرامج الحاسوبية مثل برنامج اكسل وبرنامج SPSS بالحاسب الشخصي.

المعادلات السابقة لـ a ، b هي نتائج مباشرة لطريقة تقدير المربعات الصغرى، والتي تعطي أفضل (١) مقدرات؛ نظراً لأنها غير متحيزة وأكثر دقة مقارنة بأي مجموعة أخرى. لذا، فإننا نفضل استخدام هذه الصيغ عند إجراء تحليل الانحدار للحصول على أفضل معادلات ممكنة لشرح العلاقة الخطية بين المتغيرين X ، Y.

وبمجرد استخدام صيغ المربعات الصغرى لـ a ، b ومن ثمَّ الحصول على معادلة العلاقة الخطية بين X ، Y يظهر التساؤل كيف نعرف أن هذه المعادلة تعطي معلومات أكثر حول Y بدلاً من متوسط Y نفسها مع تجاهل X؟ ويمكن الإجابة على ذلك باختبار الفروض حول مقدّرة الميل b في معادلة المربعات الصغرى الخطية. فإذا كان الميل ، b فعلياً يساوي صفر فإنه لا يوجد علاقة خطية بين X ، Y وإنما فقط يوجد خط أفقي ارتفاعه يساوي متوسط Y نظرياً لأن الحد الثاني في صيغة المربعات الصغرى لـ a سيساوي الصفر عندما $b = 0$. وهذا مكافئ للقول بأن متوسط Y هو أفضل تقدير لـ Y . ومن جهة أخرى، إذا لم تكن b تساوي صفر فإن المتغير الثابت X سوف يضيف المزيد حول شرح التغير في Y بدلاً من متوسط Y نفسها، وعليه فإن الاختبار سيكون لـ b. في الانحدار الخطي البسيط حيث يوجد متغير X واحد و b واحده فإنه يمكن استخدام واحد من الاختبارين التاليين إما اختبار t لـ b أو تحليل التباين للمعادلة الخطية. وسيتم التطرق لذلك في الأجزاء التالية.

اختبار t حول b A t Test on

تتم صياغة فرض العدم والفرض البديل لهذا الاختبار من حيث معالم المجتمع لميل الخط β . والفرض البديل يمكن أن يكون اتجاهين إذا كان لا يوجد لدينا معلومات خارجية لتحديد اتجاه العلاقة بين المتغيرين X ، Y ، أو يكون اتجاه واحد إذا كان لدينا بيانات تحدد الاتجاه (ايجابي أو سلبي) للعلاقة . ويمكن كتابة الفرض باتجاهين على النحو التالي :

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

ويتم تحديد درجه الحرية لهذا الاختبار باستخدام المعادلة رقم (11.8) حيث الرمز k يساوي عدد المتغيرات المستقلة X في المعادلة رقم (11.9) :

$$v = (n - k - 1) \qquad (11.8)$$

ونلاحظ أننا نخسر درجه حرية لتقدير a ودرجات أخرى لكل b يتم تقديره. في حالة الانحدار الخطي البسيط فإن درجة الحرية هي $(n - 2)$ لأن لدينا مقدرين هما a ، b . وبمعرفة مستوى المعنوية يمكننا إيجاد القيمة الحرجة لـ t من الجدول رقم (٨) بالملحق ومقارنتها بقيمة t المحسوبة من المعادلة رقم (11.9) :

$$t = \frac{b - \beta}{S_\beta} = \frac{b - \beta}{\sqrt{\dfrac{\dfrac{\sum (Y_i - Y_e)^2}{n - 2}}{\sum (X_i - \overline{X})^2}}} \qquad (11.9)$$

فإذا كانت قيمة t المحسوبة أكبر من القيمة الحرجة (الجدوليه) فإننا نرفض فرض العدم ونقول بأن العلاقة الخطية تم وصفها بخط الانحدار لـ Y على X . وإذا

كانت القيمة المحسوبة اقل من الحرجة (الجدوليه) فإننا لا نرفض فرض العدم وعليه لا يوجد علاقة خطيه بين المتغيرين.

تحليل التباين لخط الانحدار

Analysis of Variance for the Regression Line

تتم تجزئة مجموع المربعات ودرجات الحرية المرتبطة بالمتغير العشوائي Y وذلك لإجراء تحليل التباين لخط الانحدار . وكما ناقشنا سابقاً في تحليل التباين باتجاه واحد في الفصل التاسع ، فإننا نحصل على مجموع المربعات الكلي كمجموع لمربع الاختلافات لـ Y عن متوسطها كما هو موضح في المعادلة رقم (11.10) :

$$\text{Total } SS = \sum (Y_i - \overline{Y})^2 \qquad (11.10)$$

وتتم تجزئته إلى مجموعي مربعات- هما مجموع مربعات الانحدار SSR ومجموع مربعات الخطأ SSE. ويمكن الحصول على مجموع مربعات الانحدار بجمع مربعات الفروق بين قيمة Y المقدرة من المعادلة ، Y_e ، ومتوسط Y كما في المعادلة رقم (11.11) :

$$SSR = \sum (Y_e - \overline{Y})^2 \qquad (11.11)$$

وحساب مجموع مربعات الخطأ كمربع للانحرافات حول خط الانحدار (المعادلة رقم 11.12):

$$SSE = \sum (Y_i - Y_e)^2 \qquad (11.12)$$

ودرجات الحرية لمجموع المربعات الكلي تساوي $(n-1)$ بينما درجه الحرية لمجموع مربعات الانحدار SSR تساوي k درجه حرية أما مجموع مربعات الخطأ فإن درجه الحرية لها هي $(n-k-1)$. وبحساب مجموع المربعات يمكننا إنشاء جدول تحليل التباين كما في الجدول رقم (١١,٢) التالي :

الجدول رقم (١١.٢). جدول ANOVA للانحدار الخطي البسيط.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	F
الانحدار	*1*	*SSR*	*MRS=SSR/1*	*MSR/MSE*
الخطأ	*n-2*	*SSE*	*MSE=SSE/(n-2)*	
المجموع	*n-1*	*Total SS*		

معظم الحسابات تتم باستخدام الحاسب الشخصي والتي تعطي جدول لتحليل التباين مشابه للجدول رقم (١١,٢). ولكن إذا تطلب الأمر استخدام الآلة الحاسبة اليدوية فإنه يمكن إجراء ذلك باستخدام المعادلات الحسابية التالية. حيث مجموع المربعات الكلي يمكن حسابه كما في المعادلة رقم (11.13) ومجموع مربعات الانحدارات يمكن حسابه باستخدام واحده من المعادلات الثلاث الموضحة في المعادلة رقم (11.14) :

$$\text{Total SS} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = \sum Y_i^2 - n\overline{Y}^2 \tag{11.13}$$

$$SSR = b^2 \sum (X_i - \overline{X})^2 = b\left[\sum (X_i - \overline{X})(Y_i - \overline{Y})\right] = b\left[\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right] \tag{11.14}$$

أما مجموع مربعات الخطأ فإنه يمكن حسابه بالطرح أو استخدام الصيغة الموضحة في المعادلة رقم (11.5) :

$$SSE = \text{Total SS} - SSR = \sum Y_i^2 - a\sum Y_i - b\sum X_i Y_i \tag{11.5}$$

وتتم مقارنة قيمة F المحسوبة من جدول تحليل التباين ANOVA بقيمه F الحرجة (الجدوليه) باستخدام درجه حرية تساوي الواحد للبسط ودرجه حرية تساوي $(n - 2)$ للمقام ومستوى المعنوية المختار. فإذا كانت قيمه F المحسوبة أكبر من قيمة F الحرجة (الجدوليه) من الجدول رقم (١٠) بالملحق فإننا نرفض فرض العدم المتضمن عدم وجود علاقة خطية بين المتغيرين X ، Y وما عدا ذلك فلا يتم رفضه.

معامل التحديد The Coefficient of Determination, r^2

المقياس العددي لنسبة الاختلاف في المتغير العشوائي Y المفسر بواسطة خط الانحدار هو معامل التحديد و يكتب r^2. وفي حالة الانحدار الخطي البسيط فإنه عبارة عن مربع معامل الارتباط r والذي تم عرضه سابقاً ، ولكن r^2 يقيس كيفية تمثيل خط انحدار المربعات الصغرى للبيانات الفعلية. ولذا فإنه يحسب كنسبة لمجموع مربعات الانحدار إلى مجموع المربعات الكلي من تحليل التباين كما في المعادلة رقم (11.16) التالية :

$$\mathrm{r}^2 = \frac{\mathrm{SSR}}{\mathrm{TotalSS}} = \frac{b^2 \sum\left(X_i - \overline{X}\right)^2}{\sum Y_i^2 - n\overline{Y}^2} \tag{11.16}$$

وكذلك يمكن التعبير عن معامل التحديد بصيغه رياضية أخرى كما في المعادلة رقم (11.17) التالية :

$$\mathrm{r}^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{TotalSS}} \tag{11.17}$$

وكلما كانت قيمة r^2 كبيرة كلما دل ذلك على تمثيل أفضل لخط الانحدار للبيانات. ولكن ، في حالة أنواع معينة من البيانات مثل الأسعار والدخل وبيانات

الخصائص الاجتماعية والاقتصادية والتي تستخدم من قبل المختصين في الاقتصاد الزراعي فإنه ليس من المعتاد الحصول على قيمة لمعامل التحديد تساوي ٠,٦٠ أو قريب من ذلك والتي تعتبر جيده جداً .

وقيمة r^2 التي تعتبر جيده تعتمد بشكل كبير على نوع البيانات المدروسة. وعليه فإن هذا ملخص إحصائي معتاد يتم عرضه مع معادلة الانحدار المقدرة.

الخطأ المعياري للتقدير $S_{y \cdot x}$ The Standard Error of the Estimate

تستخدم معادلة الانحدار مبدئياً لتقدير المتغير العشوائي التابع باستخدام قيم معينة للمتغير X . وبالحصول على القيم المقدّرة للمتغير Y نرغب في معرفة مدى الاعتماد عليها. وهذا يقيس مدى قرب نقاط البيانات وانتشارها حول خط الانحدار المرسوم في شكل الانتشار. وكلما كانت تلك النقاط قريبة من الخط كلما كان الاعتماد على القيم المقدرة لـ Y باستخدام المعادلة أكثر.

والخطأ المعياري للتقدير $S_{y \cdot x}$ يعطي مقياس عددي لانتشار النقاط حول معادلة الانحدار كما في شكل الانتشار الذي يعطي تفسير مرئي أو مشاهد. وكلما كانت قيمة الخطأ المعياري للتقدير صغيره كلما كان الاعتماد أكثر على القيم المقدرة لـ Y باستخدام معادلة الانحدار ، أي كلما كانت Y_e قريبه من نقاط البيانات Y_i لقيمة X المعطاة وبذلك تكون البيانات أقل ابتعاد من خط الانحدار وعليه فإن الحالة النادرة والتي تقع فيها كل قيم Y_i على خط الانحدار فإنه لن يكون هناك اختلاف وتكون $r^2 = 1$ و $S_{y \cdot x} = 0$ ويمكن حساب الخطأ المعياري للتقدير باستخدام المعادلة الرياضية رقم (11.18) التالية :

$$S_{y \cdot x} = \sqrt{MSE} = \sqrt{(SSE)/(n-2)} \qquad (11.18)$$

حيث MSE هي متوسط مربعات الخطأ من جدول تحليل التباين ANOVA والخطأ المعياري للتقدير نظرياً مشابه للانحراف المعياري وعليه فإن تفسيره مشابه له أيضاً. فإذا كان انتشار قيم Y حول معادلة الانحدار يتوزع طبيعياً ولدينا عينة كبيرة من البيانات فإن حوالي ٦٨٪ من النقاط في شكل الانتشار سوف تكون ضمن حدود خطأ معياري واحد للتقدير أعلى وأسفل خط الانحدار ، و٩٥,٥٪ من النقاط سوف تكون في مدى خطائين معيارين للتقدير أعلى وأسفل الخط ، وتقريباً الكل، أو ٩٩,٧٪، من النقاط سوف تكون في مدى ثلاثة خطأ معياري للتقدير أعلى وأسفل الخط. وعليه فإنه يمكننا إيجاد فتره الثقة حول خط الانحدار بناء على الخطأ المعياري للتقدير وكذلك إيجاد حدود الثقة لمتوسط Y .

فترة الثقة لـــ $E(Y_e)$ Confidence Interval for

في تحليل الانحدار، ربما نهتم أكثر بتقدير متوسط الاستجابة لـ Y لقيمة معينة من المتغير X. وندرس المتوسط بسبب أن مجموعة البيانات المستخدمة لتقدير معادلة الانحدار عبارة عن عينة مختارة من عينات عشوائية كثيرة ربما تكون جمعت لقيم المتغير العشوائي Y وفقاً لمجموعه معينه من قيم X.

فإذا تم أخذ جميع العينات العشوائية المحتملة وقدّر خط الانحدار لكل واحدة، فإن الخطوط ليس من الضروري أن يكون لها نفس القاطع a والميل b. وبالتالي فإننا ربما نقدر عدة خطوط انحدار مختلفة ولكن جميع هذه الخطوط ستمر بنقاط المتوسطات متوسط X ومتوسط Y. وعليه فإن قيم X القريبة من متوسطها تكون القيم المقدرة

لـ Y المتحصل عليها من كل خطوط الانحدار الممكنة $(Y_e 's)$ وتكون أكثر تشابهاً مقارنة بحالة X البعيدة من متوسطها. ويتم الحصول على متوسط Y لقيمة معطاة لـ X من متوسط توزيع المعاينة لـ Y_e. وفي حالة أخذ عينة واحدة وحساب خط الانحدار المفرد لها بناء على بيانات العينة ، فإنه يمكننا إيجاد فتره الثقة لمتوسط Y باستخدام توزيع t كما هو موضح في المعادلة رقم (11.19) :

$$Y_e - t_{(\alpha/2,v)} \cdot S(Y_e) \le E(Y_e) \le Y_e + t_{(\alpha/2,v)} \cdot S(Y_e) \qquad (11.19)$$

حيث $S(Y_e)$ يمكن حسابها من المعادلة رقم (11.20). ويجب ملاحظة أن القيمة $S(Y_e)$ المحسوبة من المعادلة رقم (11.20) تكون كبيره كلما كانت قيم X_e بعيدة من متوسطها كما في مربع الفرق في الحد الموضح تحت الجذر.

$$S(Y_e) = S_{y.x} \sqrt{\frac{1}{n} + \frac{(X_e - \overline{X})^2}{\sum (X_i - \overline{X})^2}} \qquad (11.20)$$

مثال توضيحي An Example Problem

ترغب مديرة مبيعات محل زراعي في التنبؤ بالمبيعات الشهرية Y لشركتها باستخدام النفقات الإعلانية X. قامت بجمع بيانات للأداء السابق للمحل لمدة عشرة أشهر (الجدول رقم ١١,٣) أحسب الانحدار لـ Y على X واختبر مدى اختلاف معامل الميل عن الصفر من عدمه باستخدام مستوى معنوية ١٪ ، أحسب معامل التحديد r^2 ثم فسره ، ثم أوجد فتره ثقة ٩٥٪ لمتوسط Y عندما تكون قيمه X_e=1000 دولار.

الجدول رقم (١١.٣). النفقات الإعلانية و المبيعات الشهرية (بالألف دولار) للمحل الزراعي.

النفقات الإعلانية X	المبيعات الشهرية Y	XY	X^2	Y^2
١,٢	١٠١	١٢١,٢	١,٤٤	١٠٢٠١
٠,٨	٩٢	٧٣,٦	٠,٦٤	٨٤٦٤
١,٠	١١٠	١١٠,٠	١,٠٠	١٢١٠٠
١,٣	١٢٠	١٥٦,٠	١,٦٩	١٤٤٠٠
٠,٧	٩٠	٥٦,٠	٠,٤٩	٨١٠٠
٠,٨	٨٢	٦٥,٦	٠,٦٤	٦٧٢٤
١,٠	٩٣	٩٣,٠	١,٠٠	٨٦٤٩
٠,٦	٧٥	٤٥,٠	٠,٣٦	٥٦٢٥
٠,٩	٩١	٨١,٩	٠,٨١	٨٢٨١
١,١	١٠٥	١١٥,٥	١,٢١	١١٠٢٥
٩,٤	٩٥٩	٩٢٤,٨	٩,٢٨	٩٣٥٦٩

الحل

لحل هذا المثال نبدأ أولاً بصياغة فرض العدم والفرض البديل كالتالي:

$$H_0 : \beta = 0$$
$$H_a : \beta > 0$$

ومن ثم استخدام اختبار طرف واحد باتجاه اليمين نظراً لتوقعنا زيارة المبيعات الشهرية كلما زاد الإنفاق على الدعاية والإعلان. وسيتم اختبار الفرض أعلاه باستخدام اختبار t وتحليل التباين. ولكن أولاً سيتم حساب معالم خط الانحدار a ، b حيث نبدأ حساب b أولاً.

$$b = \frac{\Sigma X_i Y_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}} = \frac{924.8 - \frac{(9.4)(959)}{10}}{9.28 - \frac{(9.4)^2}{10}} = 52.567$$

$$a = \overline{Y} - b\overline{X} = 95.9 - (52.567)(0.94) = 46.486$$

وعليه تكون معادلة الانحدار المقدرة بطريقة المربعات الصغرى هي :

$$Y_e = 46.49 + 52.57X$$

ولاختبار الفرض لـ β باستخدام تحليل التباين ANOVA فإننا نحسب أولاً إجمالي مجموع المربعات SS :

$$\text{Total SS} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = 93{,}569 - \frac{(959)^2}{10} = 1{,}600.9$$

$$\text{SSR} = b\left[\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right] = 52.567\left[924.8 - \frac{(9.4)(959)}{10}\right] = 1{,}226.93$$

ثم نحصل على مجموع مربعات الأخطاء بالطرح كالتالي :

$$\text{SSE} = \text{TotalSS} - \text{SSR} = 1{,}600.9 - 1{,}226.93 = 373.97.$$

ويمكن الآن إنشاء جدول تحليل التباين ANOVA واختبار الفرض (الجدول رقم ١١,٤).

الجدول رقم (١١.٤). جدول ANOVA لتحليل الانحدار للنفقات الإعلانية والمبيعات الشهرية.

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	*F*
الانحدار	١	١٢٢٦.٩٣	١٢٢٦.٩٣	٢٦.٢٥
الخطأ	٨	٣٧٣.٩٧	٤٦.٧٥	
المجموع	٩	١٦٠٠.٩		

ويتضح أن قيمة F المحسوبة تساوي $F = 26.25$ في حين أن قيمة F الحرجة تساوي ١١.٢٦ من الجدول رقم (١٠) بالملحق عند مستوى ١٪ ودرجات حرية ١ ، ٨ . وحيث إن F المحسوبة أكبر من الجدوليه فإن القرار برفض فرض العدم المتضمن أن ميل خط الانحدار يساوي صفر ونقول بأن الإنفاق على الدعاية والإعلان يفسر الاختلاف في المبيعات الشهرية أكثر من متوسط المبيعات نفسها. ويمكن تحديد المقدار المفسر من ذلك الاختلاف في Y باستخدام معامل التحديد والذي يساوي :

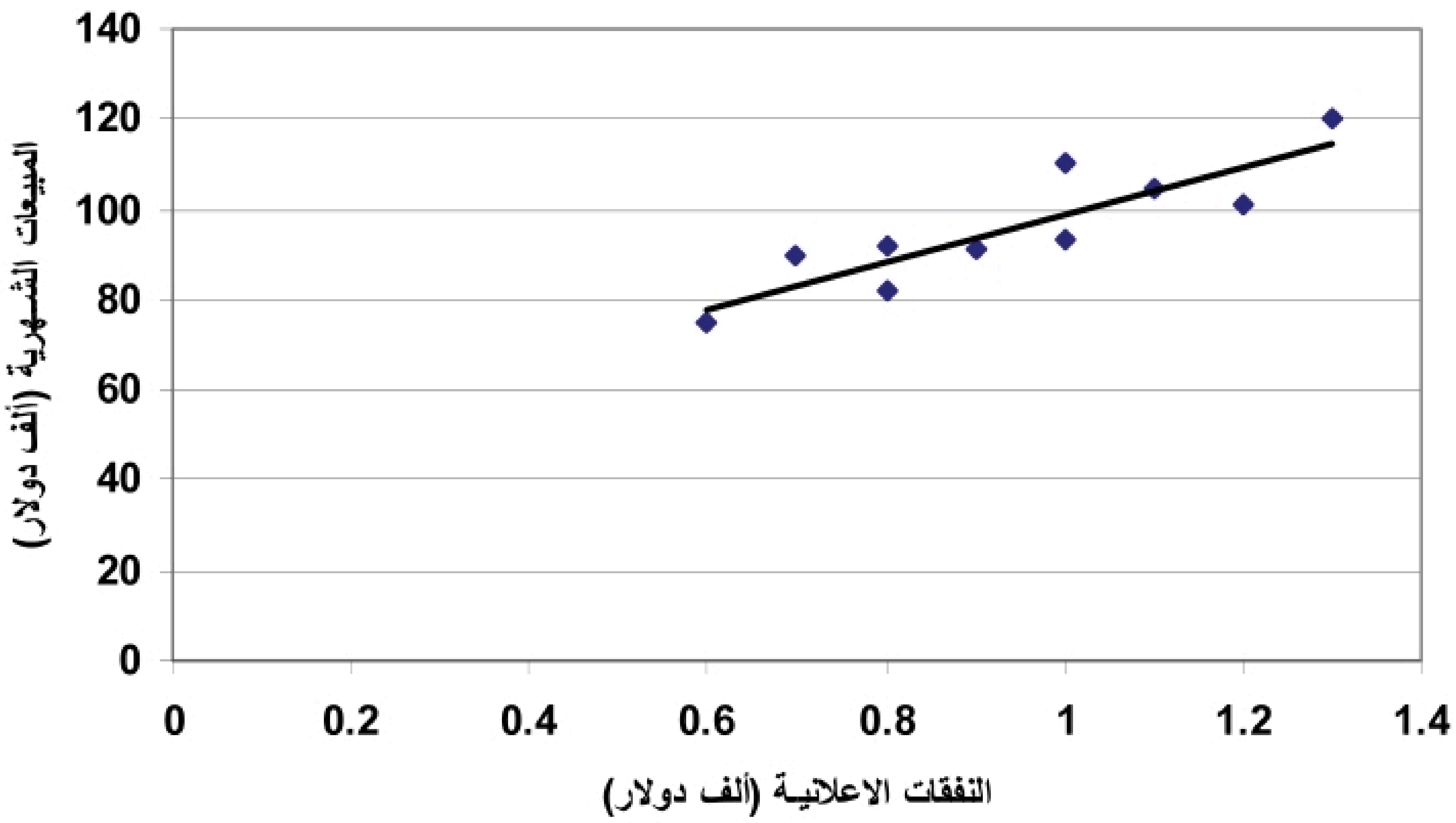
$$r^2 = \frac{SSR}{\text{Total SS}} = \frac{1{,}226.93}{1{,}600.90} = 0.77$$

وعليه فإن ٧٧٪ من الاختلافات في المبيعات الشهرية يتم تفسيرها بقيمة الإنفاق الشهري على الدعاية والإعلان. ولإجراء اختبار t لـ الميل β فإنه يجب أولاً حساب S_b ثم حساب قيمة t المحسوبة كالتالي :

$$S_b = \sqrt{\frac{MSE}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}} = \sqrt{\frac{46.75}{9.28 - \frac{(9.4)^2}{10}}} = 10.26$$

$$t = \frac{b - \beta}{S_b} = \frac{52.567 - 0}{10.26} = 5.12$$

ولاختبار الفرض فإننا نقارن t المحسوبة $t = 5.12$ بقيمه t الجدوليه والتي تساوي ٢,٨٩٦ من الجدول رقم (٨) بالملحق عند مستوى معنوية ١٪ ودرجة حرية ٨ درجات. وحيث إن قيمة t المحسوبة أكبر من قيمه t الجدوليه فإننا نرفض فرض العدم القائل بمساواة الميل للصفر (الشكل رقم ١١,٢) وهي نفس النتيجة التي تم التوصل إليها باستخدام تحليل التباين.



الشكل رقم (١١.٢). تقدير خط الإنحدار لبيانات الإعلان والمبيعات للمحل الزراعي.

ويلاحظ أن بعض الباحثين يكتبون قيمه S_b داخل قوسين وذلك تحت قيمة b عند كتابه معادلة الانحدار المقدرة، والبعض الآخر يفضل وضع قيمة t المحسوبة ووضع نجمة واحدة بجانب قيمة t إذا كانت معنوية عند مستوى ٥٪ أو نجمتين إذا كانت معنوية عند مستوى ١٪. وعند استخدام أي من الطريقتين فإنه يجب الإشارة إلى أي القيم التي تم استخدامها حتى يستطيع القارئ فهم ذلك. وعلى سبيل المثال يمكن

كتابة المعادلة المقدّرة آنفاً على النحو التالي :

$$Y_e = 46.49 + 52.57X$$
$$(5.12^{**})$$

و القيمة بين القوسين هي قيمة t المحسوبة.

ولاستخدام معادلة الانحدار لتقدير Y عندما X تساوي ١٠٠٠ دولار $X=\$1000$ فإنه يجب تحويل قيمة X أولاً حتى تتوافق مع الطريقة المستخدمة لكتابه البيانات الأصلية وبذلك تكون قيمه $X=\$1$ ثم نعوض عن قيمة X بـ ١ في المعادلة لإيجاد قيمة Y.

$$Y_e = 46.49 + 52.57X = 46.49 + 52.57(1) = 99$$

ولذا فإن المبيعات الشهرية تم تقديرها بالمعادلة وهي تساوي ٩٩٠٠٠ دولار عندما تكون النفقات الإعلانية تساوي ١٠٠٠ دولار. وهذا عبارة عن تقدير نقطة.

ولإيجاد فترة ثقة ٩٥٪ لمتوسط القيمة للمبيعات عندما تكون قيمه الإنفاق على الإعلان تساوي ١٠٠٠ دولار، يمكن استخدام صياغة فترة الثقة المشار لها سابقاً في المعادلة رقم (11.19) والمعادلة رقم (11.20) التالية :

$$Y_e - t_{(\alpha/2,v)} . S(Y_e) \le E(Y_e) \le Y_e + t_{(\alpha/2,v)} \cdot S(Y_e)$$

حيث :

$$S(Y_e) = S_{y \cdot x}\sqrt{\frac{1}{n} + \frac{(X_e - \overline{X})^2}{\sum(X_i - \overline{X})^2}}$$

ولحساب $S(Y_e)$ فأننا يجب أولاً أن نحسب $S_{y.x}$ والتي هي عبارة عن الجذر التربيعي لـ MSE كالتالي :

$$S_{y\cdot x} = \sqrt{MSE} = \sqrt{46.747} = 6.84$$

$$S(Y_e) = 6.84\sqrt{\frac{1}{10} + \frac{(1-0.94)^2}{0.444}} = 2.25$$

ومـن الجـدول رقـم (٨) بـالملحق نجـد أن قيمـة $t_{(0.025,8)}$ تـساوي 2.306 . وبالتعويض عن تلك القيم في صيغة فتره الثقة آنفاً نجد أن :

$$99 - (2.306)(2.25) \le E(Y_e) \le 99 + (2.306)(2.25)$$

$$93.81 \le E(Y_e) \le 104.19$$

وعليه فإنه عندما يكون الإنفاق الإعلاني يساوي ١٠٠٠ دولار شهرياً، فإن المتوسط الشهري الحقيقي للمبيعات يتراوح بين ٩٣٨١٠ دولار و ١٠٤١٩٠دولار بفترة ثقة ٩٥٪ وهذا الفرق أكثر من ١٠٠٠٠ دولار والتي تعتبر كبيره. وفي الواقع العملي، إذا كان مدير المنشأة يرغب في عمل دراسات من هذا النوع فإنه يجب اختيار عينه كبيرة أكثر من عشرة والتي يجب أن تقلل الخطأ المعياري للتقدير $S(Y_e)$ والذي يجعل فترة الثقة أضيق.

الانحدار عندما تكون X عشوائية Regression when X Is Random

يفترض نموذج الانحدار للمربعات الصغرى العادية أن قيم X محددة و ثابتة. لذا فإن ثقة المعالم (المقدرات) ومخاطر الأخطاء تشير إلى العينات المعادة عندما تبقي قيم X نفسها من عينه إلى عينة.

في بعض الأحيان فإنه من غير المناسب اعتبار قيم X على أنها قيم ثابتة. فمثلاً، إذا حلّلنا أسعار القمح بناء على كمية القمح فإن كمية القمح أيضاً متغير

عشوائي لأن البائع الواحد لا يستطيع التحكم في هذا السوق لأنه يوجد فيه عدد كبير من البائعين وكميته قليله ولا يستطيع التأثير على الآخرين. وعليه فإنه من غير المنطق التفكير في العينات المعادة بحيث إن كميات القمح المباعة هي نفسها من عينة لأخرى.

لذا ففي حالة أن المتغيرين X ، Y عشوائية فهل نتائجنا السابقة غير ملائمة؟ الإجابة بالنفي حيث يمكن إيضاح أن جميع النتائج السابقة هي نفسها شريطه أن:

١- التوزيعات الاحتمالية الشرطية لـ Y_i بمعلوميه X_i طبيعية ومستقلة بمتوسط $\alpha + \beta X_i$ وتباين σ^2.

٢- X_i متغيرات عشوائية مستقلة والتي لها توزيعات احتماليه $f(X_i)$ لا تحتوي على المعالم α ، β ، σ^2 ، أي أن التوزيع لا يحتوي على معالم الانحدار.

والتغير الذي يجب عمله نظراً لأن المتغير X عشوائي هو في تفسير ثقة المعالم ومخاطر الخطأ. حيث إنها تشير الآن للعينات المعادة للأزواج (X_i, Y_i) حيث كلا المتغيرين تتغير من عينة إلى أخرى. أيضاً فإن قوة الاختبار تختلف عندما تكون X عشوائية.

البواقي Residuals

الباقي e_i يعرّف على أنه الفرق بين القيمة المشاهدة لـ Y والقيمة المقدرة من خط الانحدار (المعادلة رقم 11.21):

$$e_i \equiv Y_i - Y_e \qquad (11.21)$$

ويمكن اعتباره على أنه الخطأ المشاهد كعكس للخطأ الحقيقي ε_i في نموذج الانحدار، معادلة رقم (11.22):

$$\varepsilon_i = Y_i - E(Y_i) \qquad (11.22)$$

وفي نموذج الانحدار نحن نفترض أن ε_i متغيرات عشوائية طبيعية مستقلة بمتوسط $\mu = 0$ وتباين ثابت σ^2. وعند تحديد نموذج الانحدار بصورة صحيحة للبيانات، فإننا نتوقع أن يكون الخطأ المشاهد e_i له نفس الخصائص لـ ε_i. ويمكن استخدام هذه الفكرة كقاعدة لتحليل البواقي وهي طريقة جيدة لاختبار ملائمة نموذج الانحدار. والمتوسط للبواقي e_i والتي عددها n يساوي صفر كما هو موضح في المعادلة رقم (11.23):

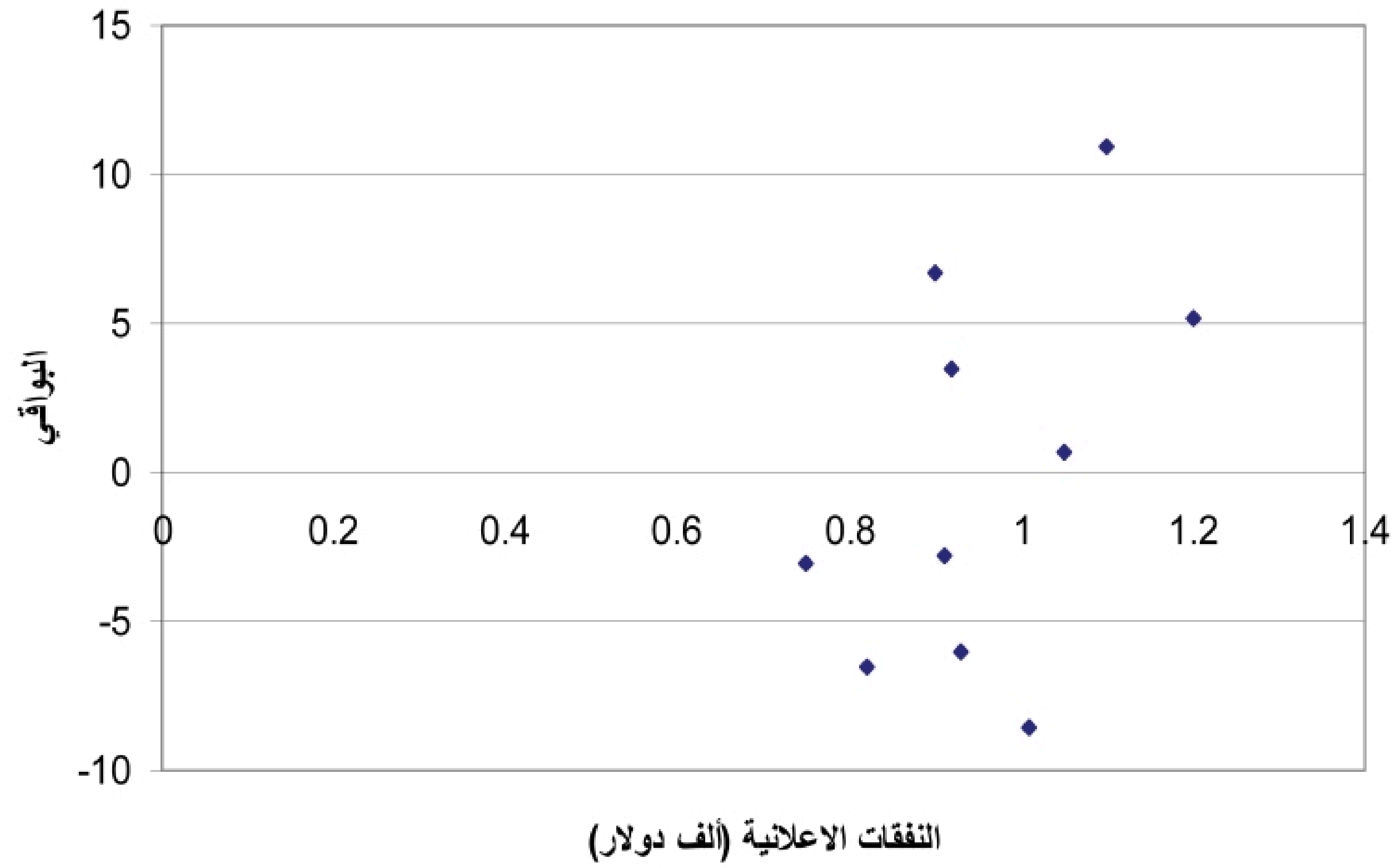
$$\bar{e} = \frac{\sum e_i}{n} = 0 \qquad (11.23)$$

وحيث إنها دائماً تساوي الصفر فإنها لا تعطي أي معلومة عما إذا كان الخطأ الحقيقي ε_i له متوسط يساوي الصفر. والتباين للبواقي e_i التي عددها n لنموذج الانحدار يساوي MSE كما هو موضح في المعادلة رقم (11.24):

$$\sigma^2 = \frac{\sum(e_i - \bar{e})^2}{n-2} = \frac{\sum e_i^{\,2}}{n-2} = \frac{SSE}{n-2} = MSE \qquad (11.24)$$

لذا فإنه إذا كان النموذج مناسب فإن المقدّر غير المتحيز لتباين الخطأ يساوي MSE.

إذا قمنا برسم البواقي على شكل انتشار مع X ، فإنه يمكننا معرفة مدى ملائمة نموذج الانحدار من شكل مجموع النقاط المتحصل عليها. فعندما يكون النموذج مقبول، فإن انتشار النقاط يتوزع بالتساوي إلى حد ما حول الصفر وتشكيل خط أفقي أو كرة بدون خط معين أو اتجاه. الشكل رقم (١١,٣) يوضح ذلك للبواقي للمثال السابق الخاص بالمبيعات والإنفاق الإعلاني.



الشكل رقم (١١.٣). شكل الانتشار للبواقي من نموذج الانحدار للنفقات الإعلانية والمبيعات الشهرية.

وفي حالة استخدام دالة انحدار غير خطية بدلاً من الخط المستقيم، فإن رسم البواقي سيأخذ شكل الخط المنحني. وفي هذه الحالة، فإننا نضيف متغير لـ X غير خطي مثل لوغارتيم X ($logX$)، أو قوى لـ X لمعادلة الانحدار كمتغير ثاني لـ X. وهذا لن يزيد من معامل التحديد فحسب بل تكون البواقي الجديدة للخط المعدّل لا تأخذ شكل المنحنى.

العامل الآخر الذي يمكن اختباره باستخدام شكل الانتشار للبواقي هو مدى ثبات التباين للخطأ. إذا أصبحت نقاط الانتشار في الاتجاه الرأسي بعيده عن الصفر واتخذت شكل مكبر الصوت أو النموذج الرباعي كلما زادت قيم X فإن تباين الخطأ كبير أو يزداد كلما زادت قيم X وليس ثابت. ويمكن أن نواجه حالات بحيث يقل التباين كلما زادت قيم X والتي تعطي شكل مشابه لشبه المنحرف بحيث تكون أكبر

للقيم الصغيرة لـ X . وكلا هذه الشروط تحدث بشكل متكرر في حالة البيانات العرضية. وفي حالة استخدام نموذج الانحدار المتعدد بأكثر من متغير X فإن رسم البواقي مقابل قيم Y_e على المحور الأفقي سوف تكون على شكل شبه منحرف إذا كان تباين الخطأ غير ثابت. نماذج المربعات الصغرى المرجحة ، والتي خارج نطاق هذا الكتاب ، تساعد على تصحيح عدم ثبات تباين الخطأ.

وعندما تكون بيانات نموذج الانحدار عبارة عن متغيرات عبر الزمن (سلاسل زمنيه) فإنه من المناسب رسم البواقي مقابل الزمن على المحور الأفقي. فإذا كان الشكل المتحصل عليه خط مستقيم بقيم سالبه للبواقي أولاً ثم قيم موجبة ، أو إذا كانت على نمط محدد بحيث يمكن رسم خط خلال البواقي تتذبذب حول الصفر فإن حد الخطأ مرتبط مع الزمن.

ويمكن استخدام اختبار ديربن- واتسون للإشارة إلى مقدار الارتباط. وحدود الخطأ غير المستقلة جديرة بالاهتمام ولكن طرق المعالجة المقترحة لها مثل استخدام المتغيرات المبطأه خارج نطاق هذا الكتاب. وإذا كانت حدود الخطأ مستقلة عن الزمن فإنه لن يكون هناك لها نمط معين وسوف تقع أو تنتشر على شكل مجموعة عشوائية أفقية حول الصفر.

أخيراً ، إذا ما قررنا إضافة متغير مستقل معين لنموذج الانحدار لزيادة القوة التفسيرية له ، فقد نرغب في رسم البواقي مقابل هذا المتغير. فإذا أوضحت البيانات اختلاف منهجي بحيث إنها في الغالب موجبه جميعها أو سالبة ، أو تأخذ نمط معين مع المتغير فإنه قد يكون من المناسب إضافة ذلك المتغير للنموذج. وفي المقابل ، إذا كانت البواقي تتوزع عشوائياً حول الصفر وتتخذ نموذج أفقي فإن إضافة ذلك المتغير للنموذج لن يضيف الكثير.

الانحدار المتعدد Multiple Regression

حيث إن نماذج الانحدار الخطي البسيط والتي تربط المتغير Y بمتغير مستقل X قويه ومفيدة، إلا أنه ربما نحتاج إلى إضافة متغيرات مستقلة X أخرى. وعموماً فإن هذا يتيح لنا شرح كامل للتغير في Y إضافة إلى تحسين قدرة النموذج التنبؤية وكذلك التقدير، وفي هذا الاتجاه يمكننا دراسة ثلاثة متغيرات أو أكثر معاً بدلاً من اثنين ولكن في هذه الحالة فإن الحسابات أكثر تعقيداً وكذلك تفسير النتائج، ويمكن تقليل الجهد الحسابي باستخدام البرامج الحاسوبية ولكن ذلك لن يساعد كثيراً في عملية التفسير.

نماذج الانحدار المتعدد Multiple Regression Models

يكون عرض نماذج الانحدار المتعدد بصورة أفضل باستخدام المصفوفات. ولكن جبر المصفوفات عموماً غير مطلوب للطلاب في العلوم الزراعية، لذا فإنه سيتم استخدام الطرق الجبرية في الشرح التالي. والطلاب الذين لديهم الرغبة في دراسة إضافية لهذا الموضوع عليهم تعلم جبر المصفوفات وأخذا مقررات متخصصة في الانحدار المتعلقة بهذا الموضوع.

وفيما يلي عرض لمعادلة الانحدار الخطي المتعدد. حيث إن متغيرات X في المعادلة يمكن أن تأخذ أي صورة، فمثلاً X_{3i} قد تكون عبارة عن حاصل ضرب X_{1i} في X_{2i} أو ربما تكون عبارة عن لوغاريثم X_2 ، أو ربما تكون مربع ، ... إلخ. في الغالب نختار نماذج متفاعلة وحدود مشروطة عليا كنماذج متكاملة ونسمي تلك النماذج الناقصة بنماذج الأثر المنخفض أو نماذج الأثر الرئيسي. وإذا كانت المعادلة خطية في معاملات الانحدار $(\beta_1, \beta_2, ...)$ فإننا نعتبرها معادلة خطية. وتستخدم β_0 كحد القاطع أو الثابت بدلاً من α وذلك لجعل العلامات أكثر تناسب (المعادلة رقم 11.25):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \quad \textit{for } i= 1, 2, \ldots, n. \qquad (11.25)$$

وافتراضات نموذج الانحدار المتعدد هي تقريباً نفس شروط النموذج الخطي البسيط وهي :

١- متوسط الخطأ العشوائي ε_i يساوي صفر وتباينه σ^2 .

٢- الأخطاء العشوائية غير مرتبطة.

٣- β_0 ، β_1 ، ..... ، β_k ، هي $(k + 1)$ معالم الانحدار و X_{1i} ، X_{2i} ، .... ، X_{ki} هي k متغيرات محدده ثابتة أو عشوائية معلومة.

٤- بهدف اختبار الفروض وتقدير فترات الثقة فإن ε_i يتوزع طبيعياً.

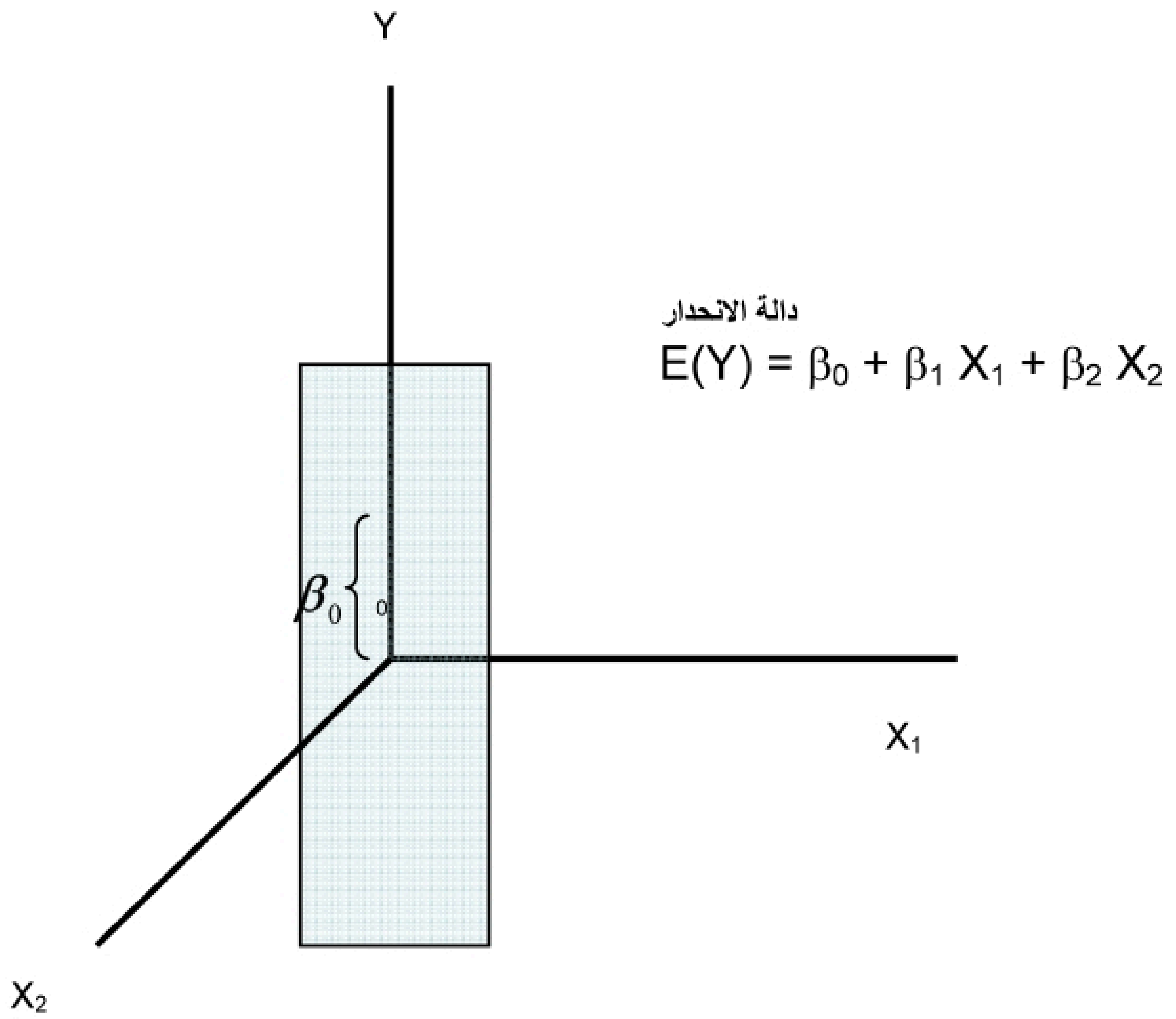
توضح الافتراضات أن متوسط Y أو القيمة المتوقعة $E(Y_i)$ لمجموعة معطاة من قيم X يمكن تعريفها بمعادلة الانحدار رقم (11.26) :

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} \qquad (11.26)$$

والتي هي عبارة عن معادلة في مستوى أبعاده $(k + 1)$ لأن Y تأخذ اتجاه مع عدد k لـ X . ولتبسيط ذلك يمكن خفض النموذج الخطي لنموذج ثلاثي الأبعاد وذلك بالتعامل مع متغيرين مستقلين X . في هذه الحالة، فإن النموذج يصف سطح المستوى كما في الشكل رقم (١١.٤). ويمكن اشتقاق المعادلات الطبيعية باستخدام المربعات الصغرى كما في الجزء الخاص بالانحدار الخطي البسيط، ماعدا أن هناك معادلة واحدة إضافية لكل X تم إضافته للنموذج. لذا ففي حالة وجود متغيرين مستقلين X فإن المعادلات الطبيعية محددة كما في المعادلة رقم (11.27) :

$$\begin{aligned}
\sum Y_i &= nb_0 + b_1 \sum X_{1i} + b_2 \sum X_{2i} \\
\sum X_{1i} Y_i &= b_0 \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i} X_{2i} \qquad (11.27) \\
\sum X_{2i} Y_i &= b_0 \sum X_{2i} + b_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^2
\end{aligned}$$

ويمكن حل تلك المعادلات آنياً لإيجاد مقدرات المربعات الصغرى b_0 ، b_1 ، b_2 لـ β_i نظراً لأن لدينا ثلاث معادلات وثلاثة مجاهيل. ولعرض طريقة المربعات الصغرى في الانحدار المتعدد نستخدم المثال التالي :



الشكل رقم (١١.٤). مسطح دالة الانحدار في حالة متغيرين مستقلين.

مثال توضيحي An Example Problem

يرغب أحد المهندسين الزراعيين في تقدير العلاقة بين إنتاجيه القطن بالرطل للوبر لكل أيكر (Y) وكمية مياه الري المستخدمة بالأنش/ايكر (X_1) وكمية السماد المستخدمة (X_2) بالرطل/ايكر وذلك لمنطقه المزارع المروية. وقد تحصل على عينه

عشوائية لعدد ١٢ مزرعة والتي تعطي البيانات الموضحة في الجدول رقم (١١,٥). المطلوب تقدير نموذج الانحدار الخطي، واختبار نموذج الانحدار ومعالم الانحدار الجزئية لمعرفة ما إذا كانت تختلف عن الصفر، ثم أوجد تقدير لمعامل التحديد مع تفسير النتائج.

لحل هذه المسألة، فإننا ندخل البيانات في برنامج حاسوبي مثل برنامج اكسل. وبالتالي نحصل على معادلة الانحدار الموضحة بالمعادلة رقم (11.28).

$$Y_e = \underset{}{248.063} + \underset{(3.88^{**})}{10.312X_{1i}} + \underset{(6.27^{**})}{4.285X_{2i}} \quad R^2 = 0.97 \quad (11.28)$$

الجدول رقم (١١.٥). بيانات الانحدار المتعدد لإنتاجية القطن وكمية مياه الري ومستوى التسميد.

إنتاجية القطن y	كمية مياه الري x_1	مستوى التسميد x_2
٤٢٠	٥	٣٠
٤٥٨	٩	٣٠
٤٠٠	٤	٢٥
٥١٠	٩	٤٠
٥٥٠	١٠	٤٠
٤٨٠	٧	٣٥
٥٧٥	١٢	٥٠
٤٦٠	٦	٣٥
٥٣٠	٩	٤٥
٦١٠	١٢	٥٥
٥٩٠	١٢	٥٥
٥٢٤	٩	٤٥

وقيمة t المحسوبة لكلا معاملات الانحدار الجزئية موضحة بين الأقواس تحت كل معامل في المعادلة رقم (11.28). وهي تختلف عن الصفر عند مستوى معنوية ١٪ ؛ نظراً لأن قيمة الاحتمال p اقل من ٠,٠١ ولذلك تم وضع نجمتين على القيمة. ويمكن إيجاد ٩٥٪ فتره ثقة لنماذج الانحدار والتي تساوي :

$$4.304 \le \beta_1 \le 16.319$$

$$2.740 \le \beta_2 \le 5.831$$

والخطأ المعياري للتقدير والذي نستخدمه لتحديد فترات الثقة تم حسابه بنفس الطريقة التي استخدمت في حاله الانحدار الخطي البسيط ولكن العلامات تختلف (معادلة رقم 11.29).

$$S_{y \cdot x_1, x_2} = \sqrt{MSE} \tag{11.29}$$

وقد أوضح تحليل التباين لنموذج الانحدار بأن مستوى الانحدار ككل له ميل يختلف عن الصفر نظراً لأن قيمة F المحسوبة $F = 164.389$ معنوية عند مستوى احتمال ١٪ (انظر ملحق هذا الفصل للإطلاع على نتائج الحاسب لهذا النموذج).

معامل التحديد المتعدد $R^2 = 0.97$ يشير إلى أن نموذج الانحدار يفسر ٩٧٪ من التغير في إنتاجية القطن. ويتم حساب R^2 للانحدار المتعدد بنفس الطريقة المستخدمة في النموذج البسيط والموضحة في المعادلة رقم (11.30).

$$R^2 = SSR/TotalSS \tag{11.30}$$

وكذلك فإن رسم البواقي هي جزء من نتائج الحاسوب. وتشير نتائج رسم البواقي مقابل كلا المتغيرين X_1 ، X_2 أن لها النمط الطبيعي. وكما ناقشنا في الجزء السابق فإن رسم البواقي هو إجراء مهم للتحقق من صحة النموذج الخطي.

ويمكن تفسير معالم الانحدار الجزئية مثلاً $b_1 = 10.312$ كالتالي. زيادة كمية المياه المستخدمة في الري بمقدار ١ بوصه/ ايكر تؤدي لزيادة إنتاجية القطن بمعدل ١٠,٣١٢ رطل للوبر/ايكر مع افتراض بقاء مستوى السماد (X_{2i}) ثابت. وإذا كانت إشارة المعالم b سالبه بدلاً من موجبه فإن زيادة X_{1i} بوحدة واحدة سيؤدي لنقص في قيمة Y_e بمقدار يساوي b_1 مع افتراض أن قيم X الأخرى ثابتة.

الارتباط الخطي المتعدد Multicollinearity

التفسير السابق لمعالم الانحدار الجزئية غير مناسب عندما تكون المتغيرات المستقلة X مرتبطة بقوة مع بعضها. وعندما تكون المتغيرات المستقلة في النموذج الخطي مرتبطة داخلياً بدرجه عاليه فأننا نقول بظهور الارتباط الخطي. في هذه الحالة من الارتباط الخطي تكون مقدرات المربعات الصغرى b_0 ، b_1 ، ... b_k لمجتمع من معالم الانحدار β_0 ، β_1 ، ... β_k غير كفؤة ؛ نظراً لأن التباينات لتوزيعات المعاينة كبيره جدا. لذا فإن b_i لا تعطي إجابة مناسبة سواء بالنسبة لقيمتها نفسها أو إشارتها ، أي أن المعرفة التي لدينا عن النماذج الخطية بناء على النظرية الخاصة بموضوعنا والتي نستخدمها لصياغة وتفسير النموذج تشير إلى أن قيم b المقدرة أو إشاراتها تقع خارج الخط. فإذا كانت درجه الارتباط الخطي حادة ، فإن أحد المتغيرات المستقلة X يمكن كتابته كمربع خطي لمجموعة جزئية من الآخر ونموذج الانحدار يحتوي فعلاً مقدرين لنفس البيانات. والحل في مثل هذه الحالة هو حذف المتغير X المخالف من النموذج ، ولكن لمعظم الحالات فإن العلاقة الخطية بين المتغيرات المستقلة X غير واضحة ونحتاج لطرق متقدمة لاستخدامها وليست من هدف موضوعنا هنا. وللحصول على معلومات إضافية عن هذا الموضوع يمكنك الرجوع لكتاب متقدم في الانحدار المتعدد أو الاقتصاد القياسي.

السلاسل الزمنية
Time series

يهتم الباحثين والمدراء ومسوقي المنتجات الزراعية بالسلاسل الزمنية لأنها تساهم في شرح ما حدث في الماضي والتنبؤ بالأحداث المستقبلية. وتحليل السلاسل الزمنية عبارة عن اختبار مبدئي للبيانات التي تم جمعها عبر الزمن مثل الأسعار الشهرية المدفوعة من قبل المزارعين للمدخلات منذ سنة ١٩٠٣. وبصفة عامة فإن السلاسل الزمنية عبارة عن مجموعه من البيانات الإحصائية والتي يتم جمعها وتسجيلها أو ملاحظتها خلال فتره زمنيه محدده. ويحدث التغير في السلاسل الزمنية نتيجة مجموعة من العوامل تؤثر عليها. فعلى سبيل المثال، في أوائل ١٩٠٠ كان المزارعين لا يشترون مدخلات للإنتاج ماعدا بعض العناصر مثل حيوانات الجر والآلات. ولكن مع تطور الزمن و توفر التسميد التجاري إضافة للبذور المهجنة والجرارات والتجهيزات الملحقة بها وجد المزارعين أنفسهم يشترون الكثير والكثير من العناصر المستخدمة في عملياتهم الزراعية. لذا أصبحت الأسعار المدفوعة للمدخلات أكثر أهمية كلما زادت قائمة تلك المدخلات. وهناك مجموعة من الحوافز لمراقبة البيانات واستخدامها كمقياس لرفاهية المزارع بالإضافة إلى البيانات الاقتصادية المهمة بعض منها على مستوى الأعمال الفردية المحلية والبعض الآخر يتم إصداره من قبل الصناعة أو المصادر الحكومية.

وإذا افترضنا أن هناك مكونات عادية ودورية تتفاعل معاً بطريقة يمكن التنبؤ بها لإعطاء السلاسل الزمنية الحالية، فإنه يمكن تحليلها بهدف التنبؤ بالسلسلة لفترة زمنيه مستقبلية. وتحليل السلاسل الزمنية طريقه تحاول فصل البيانات إلى مكوناتها التالية:

١- الاتجاه الزمني.

٢- التقلبات الدورية.

٣- التغيرات الموسمية.

٤- التحركات العرضية أو الفجائية.

الاتجاه الزمني يشير إلى الحركة الانسيابية باتجاه الأعلى أو الأسفل والتي تتميز بها السلسلة الزمنية خلال فتره من الزمن طويلة مثل عشر سنوات أو أكثر. ويعزى ذلك للتأثير الكامن للقوى الأساسية المؤثرة في البيانات مثل النمو السكاني، التغيرات التكنولوجية والتحوّل إلى السعات الكبيرة في أذواق وتفضيلات المستهلكين. لذا فإن الاتجاه يوضح حركه ثابتة ويتطلب وصفها نوع مختلف من النماذج عن تلك المستخدمة للمكونات الأخرى.

التقلبات الدورية هي التحركات في الدورات الاقتصادية والتي تعكس توسع الاقتصاد من فتره الكساد المنخفضة إلى الطفرة الاقتصادية وبالتالي انكماشها باتجاه كساد جديد. والدورة الاقتصادية تكون طويلة في العادة من سنة إلى ١٥ سنة للتحرك من قاع إلى قاع والتغيرات خلال فتره الانتعاش والانكماش. ويتم ذلك حول خط الاتجاه الزمني. وتعتمد فتره الانتعاش والانكماش على مدى تعافي الاقتصاد وتأثير الأحداث الاقتصادية على بيانات السلاسل الزمنية موضع الدراسة. وبسبب طبيعة الدورات الاقتصادية لم يتم استنباط طريقه ذات دقه عاليه للتنبؤ بالتقلبات الدورية في بيانات السلاسل الزمنية. ويتم توظيف طريقة البواقي، أي بعد استبعاد أثر الاتجاه الزمني والتغيرات الموسمية من بيانات السلسلة الزمنية، ويتبقى أثر التغيرات الدورية والعرضية، وبصفه عامه لا تنفصل هذين المكونين عن بعضها. ونظراً لأن التحركات الدورية كبيره بالنسبة للتغيرات العرضية فإن هذه الطريقة ممكنه.

التغيرات الموسمية هي دورات قصيرة الأجل داخل البيانات والتي تكتمل خلال السنة وتتكرر سنوياً. ويؤدي الطقس والعادات لحدوث التغيرات الموسمية. فمثلاً في

الزراعة فإن فتره الحصاد للمحاصيل الرئيسية تتطلب شراء أنشطه بواسطة المزارعين والتي تتكرر كل سنة في نفس الوقت تقريباً. وهذا يجعل مبيعات الإنتاج الحالي عالية في فتره معينه كل سنة يعني نمط موسمي. وبيانات السلاسل الزمنية الشهرية أو الربع سنوية عادة تحتاج لاختبارها حيال التغيرات الموسمية.

التحركات العرضية في بيانات السلاسل الزمنية لا تتبع نمط محدد لأنها تحدث بسبب تأثير الحرب ، أخطار الطقس مثل الثلج والفيضانات ، البراكين والعواصف ، والإضرابات والحوادث وغير ذلك. لذا فإنه لا يوجد لدينا طريقه يمكن بها التنبؤ بمثل هذه الأحداث ولم يتم عمل نماذج لذلك. وتبقى التغيرات العرضية في البواقي بعد إزالة العوامل الأخرى من البيانات.

تحليل الاتجاه الزمني Secular Trend Analysis

نظراً لأن بيانات السلاسل الزمنية تمتد لعدد كبير من السنوات ، فقد يكون هناك تغيرات مهمة خلال تلك الفترة نظراً لتغير السكان ، تغير مستويات الأسعار ، أو تغير في تعريفات بعض العناصر المهمة مثل تعريف المزرعة للبيانات الزراعية. وقد نحتاج لقسمه بيانات السلسلة على عدد السكان للحصول على البيانات للفرد وتحليل الاتجاه بالنسبة للفرد إذا كان التغير في السكان يؤثر على الاتجاه. وربما نرغب في قسمة بيانات النقود بمؤشر يعبّر عن التغير في القوة الشرائية للدولار عبر الزمن قبل تقدير الاتجاه الزمني. إذا كانت البيانات تتطلب العمل مع تعريف المزرعة المستخدم فإنه ربما يتطلب الأمر تعديل البيانات حسب التغير في التعريف ؛ نظراً لأن إدارة الزراعة الأمريكية تعمل تغيرات دوريه لذلك. وبعد الاطمئنان بعدم وجود أي مشكلة منهجية في البيانات يمكن القيام بتقدير الاتجاه الزمني.

الاتجاه الخطي Linear Trend

يجب أولاً تحديد ما إذا كان الاتجاه خطي. ومن الوسائل المساعدة لذلك رسم البيانات مع الزمن على المحور الأفقي X باستخدام شكل الانتشار وملاحظة نمط النقاط لها. فإذا وقعت تقريباً على خط أفقي فإنه يمكننا رسم الاتجاه الخطي للبيانات باستخدام تحليل الانحدار. وعند استخدام الانحدار فنحن عادة نرمز لمتغير الزمن بـ X ونستخدم القيم الترميزية لتقدير المعادلة بدلاً من قيم السنوات. والطريقة المعتادة أن نرمز للسنة الأولى في السلسلة بـ صفر والسنة الثانية بـ ١ وهكذا حتى نصل إلى نهاية بيانات السلسلة. وعند ذلك نستخدم الترميز المناسب لـ X في معادلة الانحدار للتنبؤ بقيمه Y.

والطريقة الأخرى البديلة هي ترميز السنة الوسطى في البيانات بالرمز صفر و السنة التالية بعد ذلك بالرقم ١ وهكذا والسنوات السابقة -١ ، -٢ ، - ٣ وهكذا حتى نصل إلى نقطة البداية للبيانات. وبهذا الإجراء فإنه مجموع قيم X يساوي صفر. أو عندما يكون لدينا أعداد زوجية من السنوات يمكن ترميز البيانات الوسطية للسلسة بالرمز ١ ، -١ للرقمين التي تقع في المنتصف واستخدام الأرقام الفردية للتحرك للأمام والوراء مثل ١ ، ٣، ٥ ... إلخ و -١ ، -٣ ، -٥ ، حتى نصل إلى نهاية البيانات. وقيم X التي تم ترميزها بهذه الطريقة يكون مجموعها مساوي للصفر. ويتم اختيار أي من خيارات الترميز أكثر ملائمة مع تعريف القارئ بنوع طريقه الترميز المستخدمة للبيانات.

الاتجاه غير الخطي Nonlinear Trends

إذا كانت السلسلة الزمنية غير خطية ، فإننا نستخدم تحليل الانحدار المتعدد. ويجب أن لا يكون الاهتمام بإيجاد المعادلة الملائمة للبيانات ولكن أيضاً الصيغة الممكن استخدامها وفقاً للطبيعة الاقتصادية للسلسلة الزمنية موضع الدراسة.

ومن تلك الصيغ المنحنى الأسى والذي يفسر السلسلة الزمنية التي تزيد أو تنقص بمعدل ثابت عبر الزمن مثل نمو السكان، مبيعات المنتج الجديد، أو انتشار الأمراض المعدية (المعادلة رقم 11.31).

$$Y = ab^x \tag{11.31}$$

ويعتمد شكل الدالة على قيم $\boldsymbol{a}$ و b . حيث $\boldsymbol{a}$ هي القاطع لـ Y ولكن المنحنى يتناقص للقيم التي تكون عندها b بين الصفر والواحد ويتزايد عندما تكون قيم b أكبر من الواحد. ويتم تقدير المعادلة بأخذ اللوغاريثم للطرفين كما في المعادلة رقم (11.32) وهذه المعادلة لوغاريثميه خطية.

$$log\ Y = log\ (ab^x) = log\ a + X\ log\ b \tag{11.32}$$

الجدول رقم (١١.٦). بيانات السلسلة الزمنية لمبيعات الأسمدة من ١٩٨٩-١٩٩٦م.

الزمن x	مبيعات الأسمدة Y	ترميز x	لوغارثيم Y
١٩٨٩	٢٦,٠	٠	١,٤١٥٠
١٩٩٠	٢٨,١	١	١,٤٤٨٧
١٩٩١	٣٠,٣	٢	١,٤٨١٤
١٩٩٢	٣٢,٨	٣	١,٥١٥٩
١٩٩٣	٣٥,٤	٤	١,٥٤٩٠
١٩٩٤	٣٨,٣	٥	١,٥٨٣٢
١٩٩٥	٤١,٢	٦	١,٦١٤٩
١٩٩٦	٤٤,٥	٧	١,٦٤٨٤

ويمكن تقديره باستخدام طريقة المربعات الصغرى. ولتقدير المعادلة يتم ترميز X كما تم شرحه في الجزء السابق لنقل من صفر إلى n لعدد n من السنوات ثم نأخذ اللوغاريتم لقيم Y. وبالتالي فإن القيم المقدرة لمعالم الانحدار هي **loga** و $logb$. ونحصل على تقدير لقيم Y ثم نأخذ اللوغاريتم العكسي لـ $(logY)$ للحصول على قيم Y المقدرّة. ويمكن إيضاح ذلك من خلال المثال التالي.

تم الحصول على المبيعات السنوية للأسمدة في المنطقة الزراعية (الجدول رقم ١١,٦) ونرغب في تقدير الاتجاه الزمني باستخدام المنحنى الأسى.

لحل هذا المثال يتم أولاً ترميز السنوات قيم X وإيجاد اللوغاريثم لقيم Y ثم تقدير المعادلة الأسية باستخدام الانحدار الخطي. وتوضح المعادلة رقم (11.33) المعادلة المقدرة حيث تشير الأرقام بين الأقواس لقيم t المحسوبة والتي يتضح معنويتها عند مستوى ١٪.

$$\underset{}{\log Y = 1.4153 + \underset{(334.6^{**})}{0.0334X}} \tag{11.33}$$

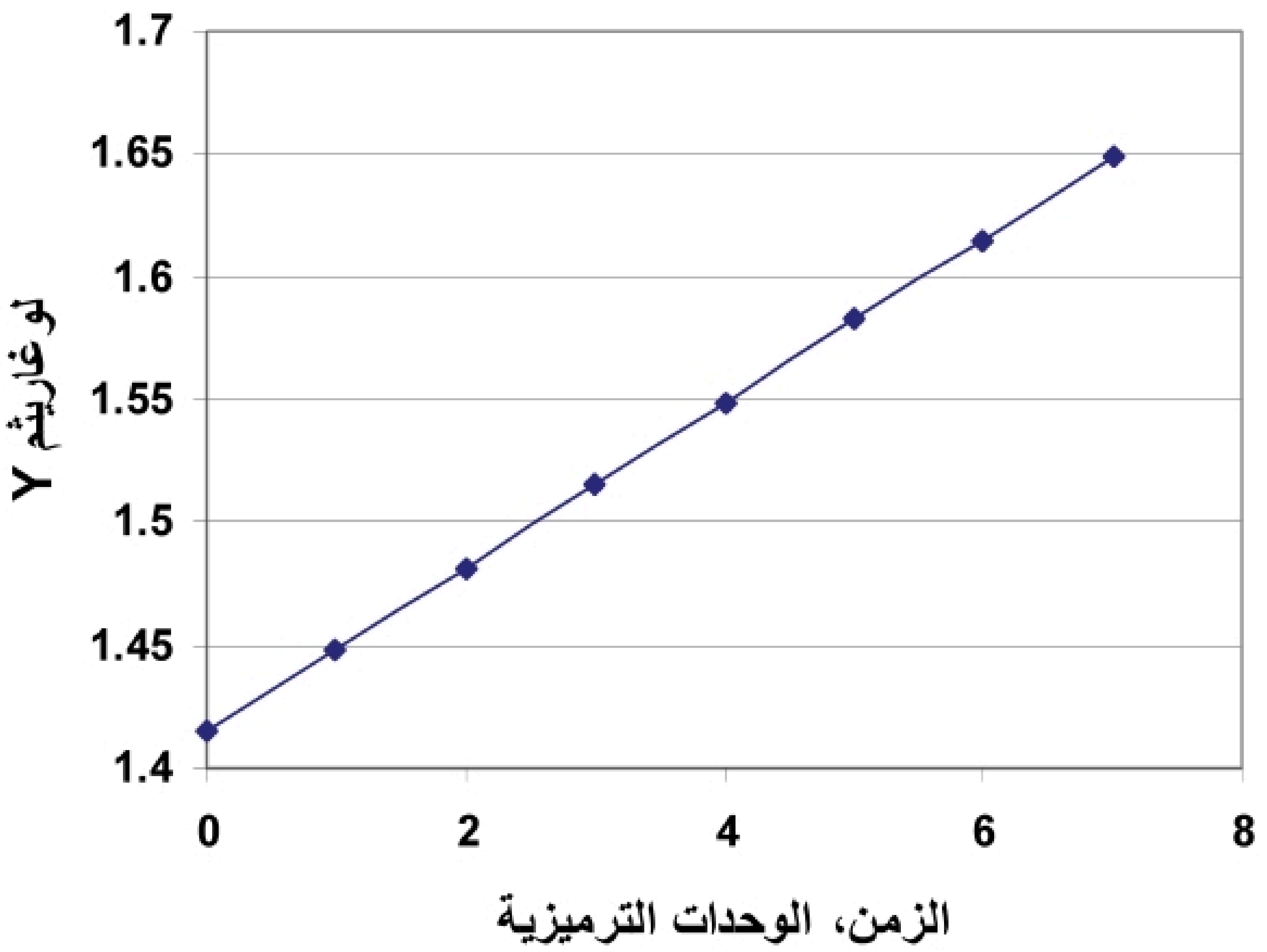
ولتقدير القيم الاتجاهية لسنه ١٩٩٧ فإننا نعوض بالترميز الخاص بالسنة ١٩٩٧ في المعادلة السابقة والحل لإيجاد قيمه $logY$ ثم إيجاد اللوغاريتم العكسي.

$$\log Y = 1.4153 + 0.0334(8) = 1.6825$$

$$\text{Antilog}(1.6825) = 48.1$$

لذا فإننا نتوقع أن تكون مبيعات الأسمدة ٤٨,١ مليون دولار لهذه المنطقة في عام ١٩٩٧ إذا كانت تتبع الاتجاه. وقد تم رسم خط الانحدار للبيانات في الشكل رقم

(١١,٥). ويجب أولاً ملاحظة أن نقاط البيانات تقريباً خطيه عندما حولّناها إلى اللوغاريثم بالنسبة لـ Y ورمّزنا قيم X وعليه فإنه ليس من المستغرب ملائمة خط الانحدار لرسم تلك البيانات.



الشكل رقم (١١.٥). تمثيل الخط للاتجاه غير الخطي.

بالإضافة لاستخدام المنحنى الأسى يمكن استخدام معادلة كثيرة الحدود من الدرجة الثانية لإيجاد الاتجاه غير الخطي للبيانات (المعادلة رقم 11.34).

$$Y = b_0 + b_1X + b_2X^2 \qquad (11.34)$$

والتي تعتبر مناسبة بصفة عامة عندما يكون هناك معدل ثابت للتغير في قيم Y عبر الزمن. ولتسهيل تحديد معدل التغير في Y يتم أولاً حساب التغير في Y من سنه إلى أخرى بالطرح. بعد ذلك يتم طرح البيانات السابقة (معدل التغير) بطرح السنة من التي بعدها. فإذا كان معدل التغير والناتج من الطرح الثاني هو تقريباً نفسه لكامل

الفترة فإن معادلة الدرجة الثانية تكون ملائمة لتقدير هذه البيانات.

ومن الصيغ غير الخطية والتي تستخدم غالباً لرسم الاتجاه دالة النمو أو الجيومتريه، المنحنى، المنحنى اللوجستي والمنحنى الأسى المعدّل ولكن لن تتم دراسة هذه الصيغ هنا.

التغيرات الموسمية Seasonal Variation

أكثر الطرق شيوعاً لتنعيم البيانات أو تقدير التغيرات الموسمية هي طريقه النسبة إلى المتوسط المتحرك. وتتطلب هذه الطريقة بأن يكون لدينا بيانات شهريه أو ربع سنوية للسلسلة الزمنية. وسيتم مناقشة الخطوات اللازمة في حالة البيانات الربع سنوية، ولكن يمكن استخدام نفس المنطق بالنسبة للبيانات الشهرية.

الخطوات هي:

١- حساب المتوسط المتحرك للأرباع الأربعة والتي تشتمل على الاتجاه الزمني والتغيرات الدورية الموجودة في البيانات الأساسية. والمتوسط المتحرك هو عبارة عن المتوسط السنوي لبيانات الأرباع الأصلية بما فيها ربع سابق. هذا يعني أننا نأخذ المتوسط للأربعة الأرباع الأولى في السلسلة الزمنية ثم نسقط الربع الأول ونضيف الربع الخامس ونحسب المتوسط للأرباع الأربعة ثم نسقط الربع الثاني ونضيف الربع السادس ونحسب المتوسط.. إلخ حتى يتم الانتهاء من كامل بيانات السلسلة.

٢- الخطوة الثانية هي قسمة البيانات الأصلية لكل ربع على متوسطها المتحرك المناظر لها للحصول على "النسبة إلى المتوسط المتحرك" والتي تحتوي الآن على المكونات الموسمية والعرضية لأن حساب المتوسط المتحرك ساهم في التخلص من الاتجاه الزمني والعوامل الدورية.

٣- نعيد ترتيب أرقام " النسبة إلى المتوسط المتحرك" للأرباع بحيث تكون أرقام الربع الأول معاً والربع الثاني معاً...إلخ ونأخذ المتوسط لهذه القيم وذلك لمحاولة تخليص البيانات من المكونات العرضية أو الفجائية. ونستخدم المتوسط المعدل لهذا الغرض، أي نسقط أعلى قيمة وأقل قيمة ثم نحسب المتوسط للقيم المتبقية. وسوف نحصل على أربعه متوسطات معدّلة متوسط للربع الأول، وآخر للربع الثاني، ...إلخ.

٤- يتم تعديل المتوسطات المعدلة بحيث يكون مجموعها مساوي لـ ٤٠٠ ومن ثم يكون المتوسط لكل منها ١٠٠. ويمثل المتوسط المعدل " مؤشر الموسمية " والذي نستخدمه للتعديل الموسمي للبيانات الأصلية وعليه تخليص بيانات السلسلة الزمنية من أثر التغيرات الموسمية. ويمكن إيضاح ذلك بالمثال التالي.

يوضح الجدول رقم (١١,٧) البيانات الخاصة بالصادرات للخضار ومنتجات الخضار من ١٩٩٢ إلى ١٩٩٦ بالبليون دولار. المطلوب حساب مؤشر الموسمية لهذه البيانات.

الجدول رقم (١١.٧). صادرات الخضار ومنتجات الخضار الربع سنوية بالبليون دولار.

السنة	والربع	صادرات الخضار	إجمالي المتوسط المتحرك الربعي	المتوسط المتحرك	نسبة البيانات الأصلية للمتوسط المتحرك
١٩٩٢	I	١,٨٦			
	II	٢,٢٠			
	III	٢,٤٢	٨,٣٩	٢,١٠	١١٥,٢
	IV	١,٩١	٨,٤٥	٢,١١	٩٠,٥
١٩٩٣	I	١,٩٢	٨,٥٩	٢,١٥	٨٩,٣
	II	٢,٣٤	٨,٦٨	٢,١٧	١٠٧,٨
	III	٢,٥١	٨,٧٥	٢,١٩	١١٤,٦
	IV	١,٩٨	٨,٦٦	٢,١٦	٩١,٧

تابع الجدول رقم (١١.٧).

السنة والربع		صادرات الخضار	إجمالي المتوسط المتحرك الربعي	المتوسط المتحرك	نسبة البيانات الأصلية للمتوسط المتحرك
١٩٩٤	I	١,٨٣	٨,٤٧	٢,١٢	٨٦,٣
	II	٢,١٥	٨,٠٦	٢,٠٢	١٠٦,٤
	III	٢,١٠	٧,٨٨	١,٩٧	١٠٦,٦
	IV	١,٨٠	٧,٨٩	١,٩٧	٩١,٤
١٩٩٥	I	١,٨٤	٧,٧٦	١,٩٤	٩٤,٨
	II	٢,٠٢	٧,٩٦	١,٩٩	١٠١,٥
	III	٢,٣٠	٨,١٢	٢,٠٣	١١٣,٣
	IV	١,٩٦	٨,٣٥	٢,٠٩	٩٣,٨
١٩٩٦	I	٢,٠٧	٨,٧٤	٢,١٨	٩٥,٠
	II	٢,٤١	٨,٧٨	٢,٢٠	١٠٩,٥
	III	٢,٣٤	٨,٧٩	٢,٢٠	١٠٦,٤
	IV	١,٩٧			

الخطوات الأولى لحساب مؤشر الموسمية هي الحصول على الإجمالي المتحرك للأرباع الأربعة لبيانات السلسلة الزمنية. لذا فإنه لبيانات ١٩٩٢ نجمع ١,٨٦ ، ٢,٢٠ ، ٢,٤٢ ، ١,٩١ للحصول على الإجمالي الذي يساوي ٨,٣٩ ونضعه قريب من متوسط البيانات مقابل للربع الثالث. نحسب بعد ذلك الإجمالي المتحرك الثاني ونضعه مقابل الربع الرابع حيث أسقطنا الربع الأول ١,٨٦ وأضفنا الربع الخامس ١,٩٢ ونستمر

بنفس الطريقة حتى نحسب جميع القيم للإجمالي المتحرك.

الخطوة التالية هي حساب المتوسط المتحرك للأرباع الأربعة بقسمه الإجمالي المتحرك لكل ربع على الرقم (٤) وكتابه الناتج في الجدول.

في الخطوة الثالثة نحسب الموسمية المحددة بقسمة البيانات الأصلية على المتوسط المتحرك ونضربها في ١٠٠ ليكون الناتج على شكل نسبة.

والآن يمكن حساب مؤشر الموسمية. ولعمل ذلك نكتب المواسم المحددة (الجدول رقم ١١،٨) وذلك للأرباع والسنوات ويكون لدينا الآن أربع قيم لكل ربع.

الجدول رقم (١١.٨). نسب البيانات الأصلية للمتوسط المحرك وفقاً للربع.

السنة	الربع I	الربع II	الربع III	الربع IV
١٩٩٢			~~١١٥،٢~~	~~٩٠،٥~~
١٩٩٣	٨٩،٣	١٠٧،٨	١١٤،٦	٩١،٧
١٩٩٤	~~٨٦،٣~~	١٠٦،٤	~~١٠٦،٦~~	٩١،٤
١٩٩٥	٩٤،٨	~~١٠١،٥~~	١١٣،٣	~~٩٣،٨~~
١٩٩٦	~~٩٥،٠~~	~~١٠٩،٥~~		
المتوسط	٩٢،٠	١٠٧،١	١١٤،٠	٩١،٦

وحيث أننا نرغب في إيجاد متوسط واحد فإننا نأخذ المتوسط للأربعة للحصول على المتوسط المعدل وذلك بشطب القيمة الصغرى والكبرى في كل عمود ثم إيجاد المتوسط لباقي القيم. وهذه المتوسطات تقريباً هي مؤشر الموسمية ولكنها غير دقيقه لأن مجموعها لا يساوي ٤٠٠. في مثالنا الحالي مجموع المتوسطات المعدلة يساوي ٤٠٤،٧. ويمكن تحويلها ليكون مجموعها ٤٠٠ بأخذ النسبة ٤٠٠/٤٠٤،٧ = ٠،٩٨٨٤ وضربها بكل متوسط للحصول على مؤشر الموسمية (الجدول رقم ١١،٩). والآن أوجدنا

مؤشر الموسمية والموضح في العمود الثالث من الجدول رقم (١١،٩)، والتي تستخدم لتخليص البيانات من أثر الموسمية.

الجدول رقم (١١،٩). حساب مؤشر الموسمية.

الربع	المتوسط المعدل	مؤشر الموسمية
I	٩٢،٠	٩٠،٩
II	١٠٧،١	١٠٥،٩
II	١١٤،٠	١١٢،٧
IV	٩١،٦	٩٠،٥

ويتم تخليص البيانات الأصلية لصادرات الخضار من أثر الموسمية بقسمه كل قيمه بمؤشر الموسمية بعد قسمته على ١٠٠ لكل ربع. فمثلاً الربع الأول من ١٩٩٢ قيمة الصادرات ١،٨٦ بليون ومؤشر الموسمية للربع الأول ٩٠،٩ (الجدول رقم ١١،٩). لذا نقسم المؤشر ٩٠،٩ على ٠،٩٠٩ ثم نقسم الرقم ١،٨٦ عليه للحصول على القيمة الخالية من الموسمية والتي تساوي ٢،٠٥ (الجدول رقم ١١،١٠). ويتم إعادة نفس العملية حتى نكمل باقي السلسلة الزمنية.

ملاحظه ختامية Endnote

(١) هذه العبارة هي نتيجة مباشرة لنظريه جاوس ماركوف والتي تنص على أنه تحت شروط النموذج $Y_i = a + bX_i + \varepsilon_i$ حيث ε_i هي الفرق $(Y_i - Y_e)$، فإن مقدرات المربعات الصغرى a ، b غير متحيزة ولها أقل تباين ضمن جميع المقدرات الخطية غير المتحيزة.

الجدول رقم (١١.١٠). بيانات صادرات الخضار ومنتجات الخضار بالبليون دولار بعد استبعاد أثر الموسمية.

السنة والربع		صادرات الخضار	مؤشر الموسمية	الصادرات بعد استبعاد الموسمية
١٩٩٢	I	١,٨٦	٩٠,٩	٢,٠٥
	II	٢,٢٠	١٠٥,٩	٢,٠٨
	III	٢,٤٢	١١٢,٧	٢,١٥
	IV	١,٩١	٩٠,٥	٢,١١
١٩٩٣	I	١,٩٢	٩٠,٩	٢,١١
	II	٢,٣٤	١٠٥,٩	٢,٢١
	III	٢,٥١	١١٢,٧	٢,٢٣
	IV	١,٩٨	٩٠,٥	٢,١٩
١٩٩٤	I	١,٨٣	٩٠,٩	٢,٠١
	II	٢,١٥	١٠٥,٩	٢,٠٣
	III	٢,١٠	١١٢,٧	١,٨٦
	IV	١,٨٠	٩٠,٥	١,٩٩
١٩٩٥	I	١,٨٤	٩٠,٩	٢,٠٢
	II	٢,٠٢	١٠٥,٩	١,٩١
	III	٢,٣٠	١١٢,٧	٢,٠٤
	IV	١,٩٦	٩٠,٥	٢,١٧
١٩٩٦	I	٢,٠٧	٩٠,٩	٢,٢٨
	II	٢,٤١	١٠٥,٩	٢,٢٨
	III	٢,٣٤	١١٢,٧	٢,٠٨
	IV	١,٩٧	٩٠,٥	٢,١٨

ملحق الفصل ١١: الارتباط والانحدار باستخدام برنامج اكسل
Appendix to Chapter 11: Correlation and Regression with Excel

برنامج اكسل يحتوي على أوامر خاصة للارتباط والانحدار. وسيتم العمل أولاً باستخدام مثال لتوضيح عمل البرنامج لإيجاد الارتباط ثم نتطرق لمثالين آخرين لإيجاد الانحدار باستخدام اكسل.

الارتباط Correlation

سنقوم باختبار نتيجة ترتيب قطع لحم البقر حسب درجة الجودة X مقابل سعر البيع للرطل Y باستخدام طريقه الارتباط من قائمه الأدوات ثم تحليل البيانات في برنامج اكسل. أولاً ندخل البيانات في ورقة العمل على شكل أعمدة منفصلة X ، Y مع مراعاة وضع الرموز في الصف الأول وإدخال البيانات ثم اختيار الأمر كما سبق. يظهر لنا جدول تلقائي نضع في الصفوف الأول مدى البيانات ثم نعلّم على خيار وجود علامات للبيانات ونحدد المكان المرغوب للنتيجة ثم نوافق على ذلك. سيظهر لنا جدول النتيجة الموضح (الجدول رقم ١١.١١).

الجدول رقم (١١.١١). نتائج الارتباط من ورقة عمل برنامج اكسل.

	درجة جودة النوع	سعر البيع للرطل
درجة جودة النوع	١	
سعر البيع للرطل	٠.٩٣٨٧٣٩٨٦	١

ويجب ملاحظة أن معامل الارتباط الناتج هو نفسه معامل الارتباط الذي سبق حسابه ماعدا التقريب (٠.٩٣٨٧ أو ٠.٩٤). ويجب ملاحظة عدم توفر اختبار للمعامل في البرنامج.

الانحدار البسيط Simple Regression

يمكن استخدام المثال الخاص بمحل بيع الأدوات الزراعية للنظر في طريقة إجراء الانحدار الخطي باستخدام برنامج اكسل. أولاً ندخل البيانات في ورقه العمل ثم نختار

الأمر انحدار Regression من قائمة الأدوات ثم تحليل البيانات. يظهر لنا جدول تلقائي ندخل مدى بيانات المتغير Y في الصندوق الأول ثم بيانات مدى المتغير X في الصندوق الثاني ونعلّم المربع الخاص بالعلامات (أسماء المتغيرات) وكذلك المربع الخاص بمستوى فتره الثقة ثم نحدد المدى للنتائج ونختار المربع الخاص بالبواقي وكذلك المربع الخاص برسم البواقي وربما المربع الخاص برسم الخط ثم نختار موافق. تظهر لنا النتائج الموضحة في الجدول رقم (١١.١٢). ويجب ملاحظة أنها نفس النتائج التي تم التوصل إليها سابقاً باستخدام المعادلات والآلة الحاسبة ماعدا التقريب.

الجدول رقم (١١.١٢). نتائج الانحدار البسيط باستخدام ورقة عمل برنامج اكسل.

خلاصة النتائج

إحصاءات الاختبار	
R المتعدد Multiple R	٠,٨٧٥٤٤٢
معامل التحديد R Square	٠,٧٦٦٣٩٨
معامل التحديد المعدلAdjusted R Square	٠,٧٣٧١٩٨
الخطاء المعياري Standard Error	٦,٨٣٧١٥
المشاهدات Observations	١٠

ANONA

	df	*SS*	*MS*	*F*	معنويةF
الانحدار	١	١٢٢٦,٩٢٧	١٢٢٦,٩٢٧	٢٦,٢٤٦٣٣	٠,٠٠٠٩٠٤
البواقي	٨	٣٧٣,٩٧٣	٤٦,٧٤٦٦٢		
الإجمالي	٩	١٦٠٠,٩			

	المعاملات	الخطـــــاء المعياري	إحصاءة t	قيمة P	الحد الأدنى ٩٥٪	الحد الأعلى ٩٥٪
القاطع	٤٦,٤٨٦٤٩	٩,٨٨٤٥٦٦	٤,٧٠٢٩٣٦	٠,٠٠١٥٣٦	٢٣,٦٩٢٦٢	٦٩,٢٨٠٣٥
X المتغير ١	٥٢,٥٦٧٥٧	١٠,٢٦٠٨٦	٥,١٢٣١١٧	٠,٠٠٠٩٠٤	٢٨,٩٠٥٩٧	٧٦,٢٢٩١٦

الانحدار المتعدد A Multiple Regression Problem

المثال الخاص في الجزء المتعلق بالانحدار المتعدد للبيانات التي تم عرضها في الجدول رقم (١١,٦) يمكن حلها بسهوله باستخدام برنامج اكسل. والإجراء المستخدم هو نفسه الذي تم استخدامه للانحدار البسيط والنتيجة المتحصل عليها من ذلك موضحه بالجدول رقم (١١,١٣).

ويجب ملاحظة أن جدول تحليل التباين يحتوي على درجتي حرية للانحدار نظراً لأن لدينا متغيرين لـ X في النموذج. ماعدا ذلك فإن النتائج مشابه في الشكل لتلك المتحصل عليها لنموذج الانحدار الخطي البسيط.

الجدول رقم (١١.١٣). نتائج الحاسب للانحدار المتعدد للنموذج الخطي لإنتاجية القطن.

خلاصة النتائج

إحصاءات الاختبار

R المتعدد Multiple R	٠,٩٨٦٥٨٨
معامل التحديد R Square	٠,٩٧٣٣٥٥
معامل التحديد المعدلAdjusted R Square	٠,٩٦٧٤٣٤
الخطاء المعياري Standard Error	١٢,٠٥٧٩٤
المشاهدات Observations	١٢

ANONA

	df	*SS*	*MS*	*F*	معنويةF
الانحدار	٢	٤٧٨٠٢,٣٧	٢٣٩٠١,١٩	١٦٤,٣٨٩٢	$٨,٢٣ \times ١٠^{-٨}$
البواقي	٩	١٣٠٨,٥٤٥	١٤٥,٣٩٣٩		
الإجمالي	١١	٤٩١١٠,٩٢			

تابع الجدول رقم (١١.١٣).

	المعاملات	الخطـــــاء المعياري	إحصاءة t	قيمة *P*	الحد الأدنى ٩٥٪	الحد الأعلى ٩٥٪
القاطع	٢٤٨,٠٦٣	١٥,٢٩٥٣٤	١٦,٢١٨٢	٥,٧٢×١٠⁻٨	٢١٣,٤٦٢٥	٢٨٢,٦٦٣٥
X المتغير ١	١٠,٣١١٦٤	٢,٦٥٥٥٥٥	٣,٨٨٣٠٤٥	٠,٠٠٣٧١٤	٤,٣٠٤٣٥٢	١٦,٣١٨٩٣
X المتغير ٢	٤,٢٨٥٤٧٧	٠,٦٨٣١٢٧	٦,٢٧٣٣٢٣	٠,٠٠٠١٤٦	٢,٧٤٠١٣٥	٥,٨٣٠٨٢

ويلاحظ من بيانات البواقي بأنها لا تحتوي على أي اتجاه معين حيث أنها تتوزع حول خط الصفر.

تابع الجدول رقم (١١.١٣).

نتائج البواقي		
المشاهدة	Y المقدرة	البواقي
١	٤٢٨,١٨٥٥	- ٨,١٨٥٥٣
٢	٤٦٩,٤٣٢١	- ١١,٤٣٢١
٣	٣٩٦,٤٤٦٥	٣,٥٥٣٤٩٤
٤	٥١٢,٢٨٦٩	- ٢,٢٨٦٨٧
٥	٥٢٢,٥٩٨٥	- ٧,٤٠١٤٩
٦	٤٧٠,٢٣٦٢	٩,٧٦٣٧٩٧
٧	٥٨٦,٠٧٦٦	- ١١,٠٧٦٦
٨	٤٥٩,٩٢٤٦	٠,٠٧٥٤٣٨
٩	٥٣٣,٧١٤٣	- ٣,٧١٤٢٦
١٠	٦٠٧,٥٠٤	٢,٤٩٦٠٤٤
١١	٥٨٦,٨٨٠٧	٣,١١٩٣٢٦
١٢	٥٣٣,٧١٤٣	- ٩,٧١٤٢٦

تمارين Exercises

١- ما الفرق بين تحليل الانحدار والارتباط في اختبار قيم X ؟ هل يمكن تحويل مسألة الارتباط إلى انحدار أو العكس؟ ناقش ذلك.

٢- ما هو " الخطأ المعياري للتقدير" ؟ هل هو معلمة أم إحصاءة ؟ ناقش ذلك.

٣- ماذا تمثل النقاط على شكل الانتشار ؟ ماذا تعني النقطة على خط الانحدار باستخدام المربعات الصغرى للعينة ؟ ناقش ذلك.

٤- البيانات التالية توضح صافي الدخل المزرعي القومي، X ، وإجمالي الأصول غير العقارية للمزرعة Y بالبليون دولار للفترة من ١٩٨٨ حتى ١٩٩٦. أحسب معامل الارتباط r ثم اختبر مدى اختلافه عن الصفر باستخدام مستوى معنوية ٥٪. ثم فسر ماذا يعني هذا المعامل في حالة هذا المثال.

السنة	X صافي الدخل المزرعي (بالبليون دولار)	Y إجمالي الأصول غير العقارية للمزرعة (بالبليون دولار)
١٩٨٨	٣٨	٢٠٤
١٩٨٩	٤٥	٢١٢
١٩٩٠	٤٤	٢٢٠
١٩٩١	٣٨	٢١٥
١٩٩٢	٤٧	٢٢٦
١٩٩٣	٤٣	٢٣١
١٩٩٤	٤٨	٢٣٧
١٩٩٥	٣٦	٢٢١
١٩٩٦	٥٢	٢٢٧

٥- تستخدم المساحة المزروعة بالايكر من الأرز X (بالمليون) للتنبؤ بالإنتاج الكلي Y (بالمليون وزن مئوي) في تحليل الانحدار. احسب خط الانحدار بالمربعات الصغرى لهذه البيانات.

X المساحة المزروعة (بالمليون)	Y الإنتاج (بالمليون وزن مئوي)
٣,٢	١٨٠
٢,٩	١٥٦
٣,٤	١٩٨
٣,١	١٧٤
٢,٨	١٧١
٣,١	١٨٠

أ) اختبر ما إذا كان معامل الميل يختلف عن الصفر باستخدام تحليل التباين ومستوى معنوية ٥٪. ثم فسر معنى الميل.

ب) احسب معامل التحديد ثم فسر معناه.

ج) تنبأ بالإنتاج عندما تكون المساحة المزروعة ٣ مليون أيكر باستخدام المعادلة لخط الانحدار. كوّن فتره ٩٥٪ لمتوسط الإنتاج المقدّر. ثم فسر معنى فتره الثقة المتحصل عليها.

د) ارسم شكل الانتشار لهذه البيانات ثم ارسم عليه خط الانحدار للمربعات الصغرى.

٦- البيانات التالية توضح العلاقة بين الدخل الشخصي للفرد X (بالألف دولار لعام ١٩٩٢) ومتوسط الاستهلاك الفردي من لحوم البقر Y (بالرطل).

X الدخل (بالألف)	Yاستهلاك لحوم البقر (بالرطل)
١٧,٦	٦٩
١٧,٨	٦٥
١٧,٩	٦٤
١٧,٨	٦٣
١٨,١	٦٢
١٨,١	٦١

أ) احسب خط الانحدار بالمربعات الصغرى للمتغير Y على X .

ب) استخدم اختبار t لتحديد ما إذا كان ميل الخط يختلف عن الصفر عند مستوى معنوية ٥٪.

ج) قدر استهلاك لحوم البقر عندما يكون الدخل الفردي ١٨ ألف دولار. أوجد فتره ثقه ٩٥٪ لهذا التقدير ثم فسر معناه.

د) احسب معامل التحديد r^2 ثم فسر معناه.

هـ) ارسم شكل الانتشار لهذه البيانات ثم ارسم خط الانحدار بالمربعات الصغرى عليه.

٧- استخدم برنامج اكسل لحل التمرين ٤.

٨- استخدم برنامج اكسل لحل التمرين ٥.

٩- استخدم برنامج اكسل لحل التمرين ٦.

١٠- تحصل مهندس زراعي على بيانات تجربة لإنتاجية القطن للايكر Y (بالرطل من الوبر) وكمية سماد النتروجين X_1 (بالرطل) وكمية مياه الري المستخدمة X_2 (انش/ايكر) . قدر خط الانحدار بالمربعات الصغرى باستخدام برنامج اكسل.

Y إنتاجية القطن (بالرطل)	X_1 النتروجين (بالرطل)	X_3 الري (انش– ايكر)
٢٥٠	٠	٠
٣٢٠	١٠	٠
٣٧٥	١٥	٣
٤٠٠	٢٠	٣
٤١٠	٢٥	٣
٤٤٠	٢٥	٦

تابع

Y إنتاجية القطن (بالرطل)	X_1 النتروجين (بالرطل)	X_3 الري (انش– ايكر)
٤٥٠	٣٠	٦
٤٧٠	٣٥	٦
٥١٠	٣٥	٩
٥٣٠	٤٠	٩
٥٣٢	٤٥	٩

أ) أوجد معادلة الانحدار ثم تحقق من جودة النموذج ككل باستخدام مستوى معنوية ٥٪. $\alpha = 0.05$.

ب) اختبر معنوية معامل الميل باستخدام مستوى معنوية ٥٪.

ج) أوجد R^2 ثم فسر معناه. ثم اوجد تقدير الخطأ المعياري.

١١- ترغب مؤسسة تصنيع زراعية صغيرة في تقدير دالة التكاليف الكلية الأسبوعية. تحصل المدير على البيانات المحاسبية من المؤسسة واكتشف العلاقة التالية :

Y إجمالي التكاليف (دولار)	X الإنتاج الأسبوعي
٤٨٠٠	٠
٤٨١٥	٥
٤٨٣٠	١٠
٤٨٥٠	١٥
٤٨٧٥	٢٠
٤٩٠٠	٢٥
٤٩٢٥	٣٠
٤٩٥٠	٣٥
٥٠٠٠	٤٠
٥٠٦٠	٤٥

قدر معادلة التكاليف الكلية باستخدام معادلة الانحدار المتعدد التي تحتوي على X ، X^2 ثم أجب على الأجزاء من أ إلى ج الواردة في التمرين ١٠ آنفاً لهذه المسألة.

١٢- ترتبط مبيعات بذور الذرة بعدد أيام الإعلان في الراديو والتلفزيون في المطاعم للمزارعين وكذلك القيمة المنفقة على الإعلان في المجتمعات، كما في الجدول التالي:

Y مبيعات بذور الذرة (١٠٠٠ دولار)	X_1 عدد الأيام	X_2 القيمة المنفقة (١٠٠٠ دولار)
٢٣٥	٨	١٦
٢٤٥	١٠	٢٠
٢٦٠	١٢	٢٥
٢٧٠	١٤	٣٠
٢٥٥	١١	٢٢
٢٦٥	١٣	٢٦
٢٧٥	١٥	٣٢
٢٤٠	٩	١٨
٢٣٣	٧	١٥

أ) اختبر ما إذا كانت معاملات الميل تختلف عن الصفر باستخدام مستوى معنوية ٥٪.

ب) حدد معامل التحديد ثم فسره.

ج) تنبأ بالمبيعات إذا استمر الإعلان في المجتمعات ١٠ أيام وكانت تكلفته ٢٤٠٠٠ دولار.

د) ما هي المشكلة التي تتوقع وجودها في هذه البيانات؟ وهل هناك أي دليل لإثبات ذلك.

١٣- مجموعه عجول أوزانها في الفئة ٤٠٠-٦٠٠رطل (x_3) تم رعيها في أماكن العشب والرعي خلال الربيع والخريف والصيف (X_1 : ١، ٢، ٣ على التوالي)، مع إعطائها حبوب حسب وزن الجسم (x_2 = نسبة وزن الجسم) وتم قياس الوزن المكتسب بالرطل Y . احسب العلاقة بين الوزن المكتسب Y والمتغيرات الأخرى مستخدماً نموذج الانحدار المتعدد.

Y الوزن المكتسب رطل/يوم	X_1 الموسم	X_2 الحبوب	X_3 وزن الجسم
٢,٥	١	١,٠	٥٩٠
٢,٠	١	١,٠	٤١٥
٢,١	١	٠,٧٥	٥٣٠
١,٠	٣	١,٠	٥٧٠
٠,٨	٣	٠,٧٥	٥٢٠
٠,٥	٣	٠,٥	٤٢٥
١,٥	٢	٠,٧٥	٥٤٠
١,٨	٢	١,٠	٤٠٥
٢,٠	٢	١,٠	٥٦٢

أ) اختبر ما إذا كانت معاملات الميل تختلف عن الصفر باستخدام مستوى معنوية ٥٪ ثم فسر تلك النتائج بناء على المسألة.

ب) احسب R^2 ثم فسر معناها.

ج) تنبأ بالوزن بالرطل لليوم لعجل وزنه ٥١٠ تمت تغذيته بحبوب مساوية لـ ١٪ من وزن جسمه وتم رعيه في فصل الصيف.

١٤- استخدم طريقه المربعات الصغرى لتقدير الاتجاه الخطي للأسعار القياسية لمنتجات المزرعة (١٩٨٢- ١٠٠) لفترة من ١٩٨٠ إلى ١٩٩٥ .

السنة	مؤشر منتجات المزرعة
١٩٨٠	١٠٢,٩
١٩٨١	١٠٥,٢
١٩٨٢	١٠٠,٠
١٩٨٣	١٠٢,٤
١٩٨٤	١٠٥,٥
١٩٨٥	١٠٦,١
١٩٨٦	١٠٥,٢
١٩٨٧	١٠٨,٣
١٩٨٨	١٠٩,٩
١٩٨٩	١٠٨,٩
١٩٩٠	١١٢,٢
١٩٩١	١١١,٧
١٩٩٢	١١٤,٦
١٩٩٣	١١٢,١
١٩٩٤	١١٦,٣
١٩٩٥	١١٧,٤

١٥- استخدم الانحدار الخطي في برنامج آكسل لتقدير خط الاتجاه للبيانات في التمرين رقم (١٤). ثم قارن كلا التقديرين. ما هو انطباعك حول هذا الاتجاه. ثم

استخدم خط الانحدار لتقدير مؤشر سعر منتجات المزرعة لعام ١٩٩٧.

١٦- يوضح الجدول التالي ديون الرهن العقاري المستحقة على الملكية الزراعية بالبليون للفترة من ١٩٨٠-١٩٩٣. ارسم خط الاتجاه للبيانات باستخدام المنحنى الاسي والخط المستقيم. حدد ما هو الأفضل لتمثيل البيانات ؟ ثم تنبأ بالديون المستحقة في ١٩٩٥ .

السنة	ديون الرهن العقاري (بالبليون دولار)
١٩٨٠	٣٠,٥
١٩٨١	٣٢,٤
١٩٨٢	٣٥,٤
١٩٨٣	٣٩,٨
١٩٨٤	٤٤,٩
١٩٨٥	٤٩,٩
١٩٨٦	٥٥,٤
١٩٨٧	٦٣,٩
١٩٨٨	٧٢,٨
١٩٨٩	٨٦,٨
١٩٩٠	٩٧,٥
١٩٩١	١٠٧,٢
١٩٩٢	١١٨,٣
١٩٩٣	١٣٣,٧

١٧- معادلة الاتجاه التالية ناتجة من رسم القطع المكافئ بتحليل الانحدار لحجم قوة العمل الزراعي في محافظه معينه.

$$Y = 49.17 + 4.23X - 0.19X^2$$
$$X = 0 \text{ in } 1975$$

X تمثل فتره ٢,٥ سنة.

Y حجم قوة العمل الزراعية بالألف.

أ) إذا افترضنا أن خط الاتجاه مثل البيانات بكفاءة، ما هو التعميم الممكن قوله حيال الطريقة التي تنمو بها قوة العمل في هذه المحافظة بالنسبة المئوية ؟ بالقيمة المطلقة؟.

ب) في الجزء (أ) افترضنا أن الخط مثّل البيانات بكفاءة. ولكن الحجم الفعلي لقوة العمل الزراعية كانت ٩٢٠٧٣ في عام ٢٠٠٠م. وهذا الرقم يشير إلى أن الخط لم يمثل البيانات جيداً. هل توافق على ذلك ؟ ناقش.

١٨- تنبأ أحد المديرين بشركتك بأن مبيعات الشركة للسنة القادمة تساوي ١٢مليون دولار باستخدام خط الاتجاه. ونتوقع عدم وجود تغيرات دوريه حادة السنة القادمة ونعتقد أن أثر الاتجاه خلال السنة سوف يكون بسيط ونتوقع استمرار النمط الموسمي في المبيعات. وذلك النمط كالتالي :

الربع	I	II	III	IV
مؤشر الموسمية	٩٠	٧٥	١٣٠	١٠٥

المطلوب التنبؤ بالمبيعات للربع الثاني والثالث للسنة القادمة.

١٩- مؤشر التغيرات الموسمية في إنتاج لحوم البقر لبعض الأشهر المختارة هو ٦٠ ليناير، ٧٠ لمارس ، ١٢٢لأغسطس ، ١٠٠لسبتمبر.

أ) في أي هذه الأشهر كان إنتاج لحوم البقر الأعلى.

ب) في أي هذه الأشهر كان الإنتاج تقريباً هو المتوسط الشهري المعتاد.

ج) زاد إنتاج لحوم الأبقار في الجنوب الغربي من ٦,٠٦ مليون رطل في يناير إلى ١١,٥٩ مليون رطل في أغسطس. ما هي نسبة التغير في الإنتاج بعد استبعاد المتغيرات الموسمية.

٢٠- مسلخ تجاري لصغار الدواجن عرض البيانات الشهرية التالية. (بالمليون رأس).

الشهر	السنة			
	١٩٩٧	١٩٩٨	١٩٩٩	٢٠٠٠
يناير	٣٣	٣٦	٤٠	٤٢
فبراير	٣٠	٣١	٣٧	٣٦
مارس	٣٥	٣٧	٣٧	٣٨
ابريل	٣٢	٣١	٣٨	٤١
مايو	٣٣	٣٩	٣٩	٤٢
يونيو	٣٥	٤١	٣٥	٣٩
يوليو	٣٤	٣٦	٤١	٤٢
أغسطس	٣٨	٣٨	٣٩	٤١
سبتمبر	٣٧	٣٦	٣٦	٤١
أكتوبر	٣٥	٣٨	٤١	٤٥
نوفمبر	٣٤	٣٥	٣٩	٣٧
ديسمبر	٣٢	٣٤	٣٧	٤٢

أ) ارسم هذه البيانات مع الزمن.

ب) احسب الاتجاه الزمني ومؤشر الموسمية الشهري باستخدام ١٢ شهر متوسط متحرك مركزه الشهر السادس.

الفصل الثاني عشر

الإحصاءات اللامعلمية

Nonparametric Statistics

معظم الاختبارات الإحصائية التي تمت مناقشتها في الفصول السابقة اعتمدت على توزيعات احتماليه خاصة أو توزيع المعاينة مثل التوزيع الطبيعي ، t ، F ... إلخ. وقد تم عمل بعض الفروض حول التوزيعات الاحتمالية للمتغيرات موضع الدراسة. وفي بعض المسائل ، قد لا نرغب في فرض مثل هذه الفروض حول معالم المجتمع وعليه فإننا نستخدم اختبارات لا تعتمد على التوزيعات أو لا معلميه.

في معظم هذه المسائل البيانات ليست كميه في طبيعتها ولكنها تكون بيانات اسميه أو ترتيبية. في حالة البيانات الاسمية يتم تسمية الأنواع وليس أكثر من ذلك. لذا فإنه في حالة الثروة الحيوانية يمكن تصنيفها على شكل عجول ، تبيعه ، تبيع ، أبقار ، ثيران ، ومجموعة الخضروات يمكن تصنيفها بالنوع مثل الهليون ، الفلفل ، الجزر ، ...إلخ. أو الناخبين يمكن تسميتهم بالأحزاب السياسية ، ... إلخ. وشكل هذه البيانات يعتمد بصفة عامه على الرقم أو التكرار الذي يقع في كل فئة. البيانات الاسمية تعبّر عن العلاقة الترتيبية وعليه يمكن استخدام الرتبة. لذا ، المستهلكين يمكن أن يرتبوا طعامهم المفضل والصحافة يمكن أن ترتب فريق كرة القدم للكلية على مستوى الوطن ، أو ربما

نرى ترتيبات موضوعية أخرى لبعض الأنواع من الخصائص الشخصية مثل الجمال أو التفضيلات. هذه الأنواع من البيانات لا تؤدي مباشرة لقياسات فعليه لأنه من غير الممكن على سبيل المثال القول كم تفضيل المستهلك للحصول على طعامه من هذا الموقع مقارنة بالموقع الثاني أو الثالث.

لذا فعند التعامل مع بيانات اسميه أو ترتيبية فإننا نستخدم الاختبارات الإحصائية الواردة في الأجزاء التالية :

اختبار مان– وتني The Mann-Whitney Test

يتم استخدام اختبار مان- وتني لتحديد ما إذا كانت العينتين المسحوبتين من نفس المجتمع أو مجتمعات مختلفة. وهذا اختبار غير معلمي مشابه لاختبار عينتين مستقلتين باستخدام اختبار t. ولكن هذا الاختبار لا يتطلب افتراض أن تتبع المجتمعات التوزيع الطبيعي كما في اختبار t.

ويعتمد اختبار مان- وتني على مفهوم إذا تم سحب عينتين عشوائيتين مستقلتين من نفس المجتمع فإن متوسط رتب العينة $r(X_{1i})$ و $r(X_{2i})$ يجب أن يكون متساوياً. وفي المقابل، إذا كان متوسط الرتب $r(X_{1i})$ من العينة الأولى لا يساوي متوسط الرتب $r(X_{2i})$ للعينة الثانية، فإننا على الأرجح لدينا عينتين من مجتمعين مختلفين ونرفض فرض العدم بتساوي متوسطات العينات.

ولعمل هذا الاختبار نأخذ بيانات كلا العينات ثم نرتبها تصاعديا على أساس أنها عينة واحدة تجميعية ونعطي كل مشاهدة رتبة ، وفي حالة القيم المتكررة نأخذ متوسط الترتيب. ثم بعد ذلك يتم أخذ مجموع الرتب للعينة الأولى ونعرّفه على أنه الإحصاءة U. و U تقريباً تتوزع طبيعياً بمتوسط $E(U)=[n_1n_2]/2$ وتباين $\sigma_u^{\ 2}=[n_1n_2(n_1+n_2+1)]/12$ وبذلك نحسب قيمة z باستخدام المعادلة رقم

(12.1) ونرفض فرض العدم إذا كانت قيمة z المحسوبة تزيد عن قيمه Z_α الجدوليه المتحصل عليها من الجدول رقم (٧) بالملحق كأي اختبار لـ Z.

$$Z = \frac{u - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \tag{12.1}$$

وتكون قيمة Z المحسوبة سالبة إذا كانت U صغيره، وهذا يحدث عندما تكون البيانات في العينة الأولى أقل من بيانات العينة الثانية؛ نظراً لأنها عبارة عن حاصل الجمع للرتب الأقل. وعلى العكس فإن قيمة Z المحسوبة تكون موجبه عندما تكون U تمثل مجموع الرتب العليا وهذا يحدث عندما تكون البيانات للعينة الأولى أكبر. ويمكن إجراء هذا الاختبار في حالة اتجاه واحد أو اتجاهين للفرض البديل. ويمكن عرض ذلك من خلال المثال التالي:

يقوم عاملان بفحص نسبة الرطوبة في شاحنة حبوب ذرة تم إحضارها للمخزن المحلي. قام العامل الأول بأخذ عشر عينات في الصباح بينما قام العامل الثاني بأخذ أحد عشرة عينة.

الجدول رقم (١٢.١). قراءات الرطوبة للعاملين.

العامل ١		العامل ٢	
١٤	٨	٦	١١
١٢	٥	٧	١٠
١٥	٤	٨	٩
١٣	٦	٦	١٢
٩	١٦	٤	٧
		٥	

يوضح الجدول قراءة نسبة الرطوبة لكل عامل. المطلوب هل هناك اختلاف في قراءات الرطوبة بين العاملين عند مستوى معنوية ٥٪.

أولاً نكتب فرض العدم والفرض البديل كالتالي :

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

وحيث إن هذا عبارة عن اختبار اتجاهين تم الحصول على قيم Z الحرجة (الجدوليه) من جدول رقم (٧) بالملحق وكانت ±١,٩٦ عند مستوى معنوية ٥٪. بعد ذلك يتم إيجاد الرتب للبيانات التجميعية وتحديد ذلك لبيانات العينة الأولى باستخدام البنط العريض (الجدول رقم١٢,٢) وذلك لحساب إحصاءة U.

$$U = \Sigma r(X_{1i}) = 1.5 + 3.5 + 6 + 10.5 + 12.5 + 16.5 + 18 + 19 + 20 + 21 = 128.5$$

و $E(u) = [n_1 n_2] / 2 = [(10)(11)] / 2 = 110/2 = 55$

$$\sigma_u^2 = [n_1 n_2 (n_1 + n_2 + 1)]/12 = [10 \cdot 11(10 + 11 + 1)]/12 =$$
$$[110(22)]/12 = 2420/12 = 201.66.$$

لذا فإن Z المحسوبة يمكن إيجادها كالتالي :

$$Z = (u - E(u)/\sigma_u = (128.5 - 55)/\sqrt{201.66} = 73.5/14.2 = 5.18$$

وبمقارنتها بقيمة Z الجدولية نجد أنها تقع في المنطقة الحرجة. وعليه فإننا نرفض فرض العدم بتساوي متوسط قراءات الرطوبة لكلا العاملين ونختم القول بأن قراءة العامل الأول أكبر.

الجدول رقم (١٢.٢). بيانات ترتيب قراءات الرطوبة.

القياس	٤	٤	٥	٥	٦	٦	٦	٧	٧	٨	٨
الرتبة	١,٥	١,٥	٣,٥	٣,٥	٦	٦	٦	٨,٥	٨,٥	١٠,٥	١٠,٥
القياس	٩	٩	١٠	١١	١٢	١٢	١٣	١٤	١٥	١٦	
الرتبة	١٢,٥	١٢,٥	١٤	١٥	١٦,٥	١٦,٥	١٨	١٩	٢٠	٢١	

اختبار الإشارة The Sign Test

اختبار الإشارة اختبار بسيط الإجراء وهو اختبار لا معلمي مكافئ لاختبار t المزدوج لعينتين غير مستقلتين. وهناك اختبارات أخرى مثل اختبار إشارة الرتب ويلكوكسن Wilcoxon ، والذي يعتبر مناسباً في مثل هذه الحالة. في اختبار الإشارة نحن نفترض عادة بيانات ترتيبية وعليه نقرر ما إذا كانت المشاهدة الأولى من كل زوج أكبر من الثانية فإننا نستبدلها بإشارة موجبه وإذا كانت أقل منها نستبدلها بإشارة سالبة. واختبار فرض العدم لاختبار الإشارة عادة لعينتين مأخوذة من مجتمعين لهما نفس الوسيط لذا فإن الاحتمال للإشارة الموجبة أو السالبة مساوٍ لـ ٠,٥ لكل زوج من البيانات. لذا فإن التوزيع ذي الحدين وتقريب التوزيع الطبيعي لتوزيع ذي الحدين تعتبر توزيعات معاينة مناسبة لهذا الاختبار. ويتم توظيف أو استخدام التوزيع الثاني إذا كان لدينا عينات كبيره. ونعمل الاختبار للعينات الكبيرة $(n > 20)$ باستخدام Z المحسوبة والمعرفة كما في المعادلة رقم (12.2) حيث R عبارة عن عدد الإشارات الموجبة و n تمثل عدد الأزواج للبيانات. ونقارن قيمه Z المحسوبة باستخدام المعادلة رقم (12.2) بقيمة Z الجدولية لقبول أو رفض فرض العدم.

$$Z = (2R - n)/\sqrt{n} \qquad (12.2)$$

ويمكن إيضاح ذلك من خلال المثال التالي :

نفترض أنه تم سؤال مجموعة من ٢٢ شخصاً لتقيم منتجين غذائية جديدة باستخدام معيار من ١ - ٢٠ حيث يكون ٢٠ هو الأفضل. والبيانات موضحه في الجدول رقم (١٢،٣) التالي:

الجدول رقم (١٢.٣). ترتيب المستهلك لنوعين جديدة من منتجات الأغذية.

المستهلك	الطماطم الفرنسية المقلية	وجبة خفيفة من المشمش المجفف	الإشارة
١	١٥	١٨	-
٢	١٢	٨	+
٣	٤	١٨	-
٤	١٩	١١	+
٥	١٠	٥	+
٦	١٧	١٩	-
٧	١٣	١٦	-
٨	٩	١٥	-
٩	١٤	١٢	+
١٠	١٦	١٨	+
١١	١١	١٣	-
١٢	١٥	١٨	-
١٣	٨	١٢	-
١٤	١٢	١٧	-
١٥	١٦	١٩	-
١٦	١٧	١٥	+
١٨	١٣	١٥	-
١٩	١٨	١٥	+
٢٠	١٥	١٧	-
٢١	١٦	١٩	-
٢٢	١١	١٧	-

اختبر ما إذا كان وسيط الرتب لكلا المنتجين متساوي عند مستوى معنوية ٥٪.

نكتب فرض العدم والفرض البديل كالتالي :

$$H_0 : p = 0.5$$

$$H_a : p \neq 0.5$$

وحيث إن هذا اختباراً ذا اتجاهين ؛ لأنه لم يسأل هل أحد المنتجين أفضل من الآخر أم لا. عدد الإشارات الموجبة لهذه المجموعة من البيانات يساوي ٧. ولذلك نستخدم المعادلة رقم (12.2) لإيجاد قيمة Z ثم نقارنها بـ Z الحرجة (الجدوليه) من الجدول رقم (٧) بالملحق والتي تساوي ±١,٩٦ لنقرر بقبول أو رفض فرض العدم ، وحيث إن قيمة Z المحسوبة تساوي -١,٧١ لذا فإننا لا نستطيع رفض فرض العدم.

$$Z = (2R - n)/\sqrt{n} = (2(7) - 22)/\sqrt{22} = (14 - 22)/4.69 = -1.71$$

وبذلك نقول بأنه ليس هناك تفضيل من المستهلكين لأحد المنتجات على الآخر.

اختبار التتابع The Runs Test

في الجزء السابق في اختبار الإشارة استخدمنا عدد الإشارات الموجبة ، ولكن ربما يكون اهتمامنا في عدد تتابع الإشارات الموجبة والسالبة بحيث نفرق التتابع بتسلسل الأرقام بنفس الإشارة. لذا إذا كان لدينا التسلسل التالي :

$$- \quad - \quad - \quad - + + + +$$

نلاحظ أنها تحتوي على تتابعين الأول بعدد ٤ إشارات سالبة والثاني بعدد أربع إشارات موجبه. وفي المقابل فإن التسلسل :

$$- + \quad - \quad - + + \quad - + + +$$

يحتوي على عدد ٦ تتابعات، الأول بعدد إشارة سالبة واحدة، والثاني بإشارة موجبة واحدة والثالث بإشارتين سالبتين والرابع بإشارتين موجبتين والخامس بإشارة سالبه واحدة والسادس بثلاث إشارات موجبة. وعدد مرات التتابع في التسلسل يمكن أن يتراوح من واحد (جميع الإشارات هي نفسها) إلى n (الإشارات متناوبة). ويمكن اختبار فرض العدم العشوائية بتوقع عدد مرات التتابع لعرض التسلسل العشوائي للأرقام في مجموعة البيانات. ويتم رفض فرض العدم إذا كان هناك عدد تتابعات قليل جداً أو كثير جداً في التسلسل.

ولإجراء الاختبار، نحدد الوسيط لمجموعة البيانات ثم نشير للمشاهدات تحت الوسيط بإشارات سالبة والتي تلي الوسيط بإشارات موجبة. ولذلك يكون عدد المشاهدات للإشارات السالبة n_1 مساوي لعدد المشاهدات للإشارات الموجبة n_2 ؛ نظراً لحسابنا الوسيط. بعد ذلك نحسب عدد التتابعات r في البيانات ثم نقارن r بقيمة r الجدوليه من الجدول رقم (١٣) بالملحق لاتخاذ القرار حيال فرض العدم برفضه أم لا.

ويمكن استخدام اختبار التتابع لفحص عشوائية حد الخطأ في نموذج الانحدار وهناك مثال معطى لهذا التطبيق في التمارين في نهاية الفصل.

ويعبر المثال التالي والمتضمن بأن التسلسل التالي عبارة عن مجموعة عشوائية من الأرقام ذات القيم الأقل من ٥٠. استخدم اختبار التتابع لبحث ما إذا كانت هذه البيانات عشوائية بمستوى معنوية ٥٪ والبيانات هي :

15	6	29	48	32	12	17	39	31	4
7	16	45	49	13	20	2	25	28	33

ولحل هذا المثال، نحسب الوسيط ثم نكتب عدد الإشارات الموجبة والإشارات السالبة لتحديد القيم قبل وبعد الوسيط ثم نحدد عدد مرات التتابع. الوسيط يساوي ٢٢,٥ ونكتب الإشارات الموجبة كبديل للقيم الأكبر من الوسيط والإشارات السالبة كبديل للقيم الأقل من الوسيط وتكون الإشارات المتحصل عليها كالتالي:

- -+++ - -++ - - -++ - - -+++

ويكون العدد الكلي للتتابع من هذا الاختبار $r = 8$. ومن الجدول رقم (١٣) بالملحق حيث $n_1 = n_2 = 10$ نجد أن قيمه r الحرجة تساوي ٦ أو أقل.

لذا فإننا لا نستطيع رفض فرض العدم للعشوائية عند مستوى معنوية ٥٪

ارتباط الرتب لسبيرمان Spearman's Rank Correlation

تحليل الارتباط الذي تم بحثه سابقاً ينسب لبيرسون والمبني على افتراض قياس المتغيرات X و Y باستخدام مقياس فتري. ولكن كما ناقشنا سابقاً، قياس الفترة قد يكون غير ممكن أو غير مناسب في بعض الحالات أو الظروف. وهو أيضاً يفترض أن X ، Y تتبع توزيعين طبيعيين. معامل ارتباط الرتب لسبييرمان r_s ، قد يكون هو الأنسب والمستعمل بكثرة كمقياس لا معلمي يقيس الارتباط ولا يتطلب الشروط المذكورة أعلاه. ويتم حساب الارتباط عادة باستخدام الرتب للبيانات وليس المشاهدات الأصلية.

وتفسير $r_s = +1$ يعني أن العينتين لها نفس الرتب بينما عندما $r_s = -1$ يعني أن العينتين لها نفس الرتب عكسياً وقيم r_s بين -١ و +١ تشير لدرجه ارتباط أقل من الارتباط التام.

ولحساب r_s فإننا أولاً نرتب البيانات في كل عينه من الأصغر إلى الأكبر. ثم نأخذ الفرق لكل زوج من الرتب d_i ثم نربعه ونوجد المجموع الكلي للمربعات ونعوض بالقيم في المعادلة رقم (12.3).

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{12.3}$$

ولإيضاح ذلك نفترض المثال التالي الخاص بأصناف الجريب فروت. حيث تم اختيار عينه من الأسر (الزوج والزوجة) وتم سؤالهم لتقيم المنتج من ١ إلى ٢٠ بحيث يكون رقم ٢٠ هو الأفضل. والمطلوب تحديد ما إذا كان الترتيب مرتبط باستخدام مستوى معنوية ٥٪ ؟ للبيانات المعطاة في الجدول التالي رقم (١٢,٤).

الجدول رقم (١٢.٤). ترتيب تفضيلات الزوج والزوجة لمذاق الجريب فروت الأبيض.

الأسرة	تفضيلات الزوج X	تفضيلات الزوجة Y	رتب X	رتب Y	d_i
١	١٢	١٤	٤,٥	٦	١,٥ -
٢	١٧	١٣	١٠	٤,٥	٥,٥
٣	٩	١٢	١	٣	٢ -
٤	١٨	٢٠	١١	١١	٠
٥	١٥	١٧	٨	٩	١ -
٦	١٤	١٠	٧	١	٦
٧	١٠	١١	٢	٢	٠
٨	١٦	١٣	٩	٤,٥	٤,٥
٩	١٣	١٥	٦	٧	١ -
١٠	١١	١٨	٣	١٠	٧ -
١١	١٢	١٦	٤,٥	٨	٣,٥ -

$$r_S = 1 - \frac{6[-1.5^2 + 5.5^2 + (-2)^2 + 0^2 + (-1)^2 + 6^2 + 0^2 + 4.5^2 + (-1)^2 + (-7)^2 + (-3.5)^2]}{11(11^2 - 1)}$$

$$r = 1 - \frac{936}{1320} = 0.291$$

ونقارن r_s المحسوبة r_s = ٠,٢٩١ والقيمة الحرجة لـ رو *rho* والمتحصل عليها من الجدول رقم (١٤) بالملحق لعدد n = 11 والتي تساوي ٠,٦٢٣ أو أكبر.

وحيث إن قيمه r_s المحسوبة أقل فإننا لا نستطيع رفض فرض العدم المتضمن أن (رو*rho*) تساوي الصفر. وعليه بالنسبة لمثالنا يمكن القول بأنه لايوجد هناك ارتباط بين ترتيب الأزواج والزوجات لتفضيلاتهم.

تمارين Exercises

١- قام الطبيب البيطري بمعاملة مجموعتين من العجول بمطهر وذلك باستخدام جرعتين مختلفة. حيث تم إعطاء المجموعة الأولى جرعة التعقيم المعيارية والمجموعة الثانية عقار جديد صنع خصيصاً لهذا الغرض. وتم تسجيل عدد الساعات اللازمة حتى تختفي الأعراض على النحو التالي:

الجرعة القياسية ١٥ ، ٩ ، ١٢ ، ٢٢ ، ١٤ ، ٩ ، ١٠ ، ١٣

العقار الجديد ٧ ، ٨ ، ١١ ، ٦ ، ٧ ، ٧ ، ٤ ، ١١ ، ١٣ ، ٥

المطلوب هل العقار الجديد أكثر فاعلية ؟ استخدم اختبار مان- وتني U عند مستوى معنوية ٥٪.

٢- ترغب تعاونيه الحبوب المحلية في اختبار أسرع خطوط الشاحنات للنقل إلى حالة

التخزين، والمخدومة بخطوط الشحن لشركتي رودهوق Roudheg وفرايت- آر – يو اس Freights-R-US. تمت مراقبة العينتين لتحديد وقت النقل لصالة التخزين بالساعات وتسجيل النتائج التالية:

شاحنات Roudheg ١٠، ٨ ، ١٣ ، ١٥ ، ١٧

شاحنات Freights-R-US ١٤ ، ٩ ، ١٢ ، ١٤ ، ١١ ، ٨ ، ١٣ ، ٢٠ ، ١٦

باستخدام اختبار مان- وتني U اختبر ما إذا كان هناك اختلاف في أداء خطي الشحن عند مستوى معنوية ٥٪ .

٣- مصنع لتغليف الخوخ يضع أنصاف الخوخ في علب سعه ١٦ أونصة. قام مدير المصنع بمعايرة الآلة لوضع متوسط ١٦,٠٥ أونصه في العلب. وقام موظفو التحكم بالجودة بسحب ١٥ علبه كل ساعة لتحديد ما إذا كانت عمليات التعبئة سليمة. وفي الساعة الأخيرة تم اختيار مجموعة من العلب ووزن كل علبه على حدة فإذا كان وزنها أكثر من الوزن المتوسط علّمت بـ H وإذا كان أقل من المتوسط علمت بالرمز L استخدم اختبار التتابع للعينة عند متوسط معنوية ٥٪ اختبر ما إذا كان التتابع عشوائياً.

H L H L L L H H L H H L H L H

٤- يرغب أحد معلمي العلوم الزراعية في الثانوية العامة في اختبار مدى معرفة الطلاب بطريقتي تعليم للحكم على الثروة الحيوانية. الطريقة الأولى هي الإجراء القياسي وذلك بتعليم الطالب في المختبر باستخدام حيوانات حية. والطريقة الثانية هي استخدام سلسلة من مقاطع الفيديو المعدة من قبل خبير متخصص. قام بتقسيم طلاب الفصل لمجموعتين ووزع الطلاب بحيث يكون كل طالب في المجموعة له نظير في المجموعة الأخرى مشابه له في المهارات والموهبة. واستخدم التعليم القياسي

للمجموعة الأولى وتعليم الفيديو للمجموعة الثانية وبعد الانتهاء من الموضوع أعطي الطلاب نفس الاختبار. وكانت نتائجهم كما في الجدول التالي :

القياس	الفيديو
٦٦	٦٨
٦٥	٦٧
٧١	٧٤
٧٢	٦٦
٨٣	٨٠
٧٤	٧٧
٩٢	٩٦
٨١	٩٠
٨٣	٨٠
٧٦	٧٢
٨٧	٨٨
٩٥	٩٥
٩٠	٩٨
٨٥	٩٥
٨٩	٩٦
٧٥	٧٣
٩٣	٩٦
٧٩	٨٨
٨٨	٩٢
٨٢	٧٨
٧٣	٨٥

باستخدام اختبار الإشارة وعند مستوى معنوية ٥٪ هل التعليم باستخدام الفيديو أفضل.

٥- الجدول التالي يوضح البواقي لمسألة الانحدار المتعدد في التمرين رقم (١٣) للفصل الحادي عشر.

نتائج البواقي

المشاهدة	بذور الذرة المتوقعة	البواقي
١	٢٣٦,١٤١٧٩٤٥	- ١,١٤١٧٩
٢	٢٤٧,١٩٠٦٠٩	- ٢,١٩٠٦١
٣	٢٥٨,٨٣٣٥٦٥٨	١,١٦٦٤٣٤
٤	٢٧٠,٤٧٦٥٢٢٥	- ٠,٤٧٦٥٢
٥	٢٥٢,٧١٥٠١٦٣	٢,٢٨٤٩٨٤
٦	٢٦٣,٧٦٣٨٣٠٨	١,٢٣٦١٦٩
٧	٢٧٦,٠٠٠٩٢٩٨	- ١,٠٠٠٩٣
٨	٢٤١,٦٦٦٢٠١٨	- ١,٦٦٦٢
٩	٢٣١,٢١١٥٢٩٥	١,٧٨٨٤٧

باستخدام اختبار التتابع عند مستوى معنوية ٥٪ اختبر ما إذا كانت الإشارات السالبة والموجبة تتوزع عشوائياً.

٦- البيانات التالية توضح البيانات التاريخية لأسعار القهوة والشاهي.

استخدم طريقه ارتباط الرتب لسبيرمان لتحديد ما إذا كان هناك ارتباط بين الأسعار من عدمه مستخدماً مستوى معنوية ٥٪.

أسعار القهوة	اسعار الشاهي
٢٠	٣٢
١٧	٣١
١٤	٢٧
١٥	٣٠
١٨	٣٢
١٩	٣٣
١٦	٣٠

٧- افحص مؤشرات السعر والإنتاج التالية لتحديد ما إذا كان هناك بينهما ارتباط مستخدماً طريقه ارتباط الرتب لسبيرمان عند مستوى معنوية ٥٪ $\alpha = 0.05$.

مؤشر الإنتاج	مؤشر السعر
٩٩	١٤٩
٨٨	١٢٨
١٠٢	٩٠
٩٦	٦٨
٩٦	٧٢
٩٣	٩٠
٩١	١٠٩
٩٤	١١٤
١٠٦	١٢٢
١٠٣	٩٧

الفصل الثالث عشر

الأرقام القياسية

Index Numbers

في الفصل السابق تم استخدام الأرقام القياسية لعرض التغيرات الموسمية في بيانات السلاسل الزمنية. ونهدف الآن لتطوير مقياس يساعد في تلخيص خصائص القيم الكبيرة من البيانات. يمكن استخدام الأرقام القياسية لهذا الغرض كونها تربط قيم المتغير التابع مثل المبيعات أو الأسعار بالمتغير المستقل مثل الزمن أو الدخل. وبالرغم من استخدام الأرقام القياسية في مواضيع عديدة في الاقتصاد إلا أن التطبيقات الطبيعية لها تكمن في وصف الأنشطة الاقتصادية والتجارية مثل التغير في الأسعار، الإنتاج، الأجور والتوظيف خلال فتره من الزمن. ويوجد هناك نوعان من الأرقام القياسية بسيطة ومركبة. يعبّر الرقم القياسي البسيط عن العلاقة بين عددين بحيث يستخدم أحدهم كأساس. وقد تكون الأرقام القياسية البسيطة على شكل نسبة مئوية وتستخدم لإيضاح التغير النسبي في سعر العمل من فتره زمنية لأخرى. أما الأرقام القياسية المركبة فتجمع معلومات من مجموعات متعددة لأسعار المزارعين يشمل مجموعة من أسعار المنتجات الزراعية في رقم قياس واحد يوضح كيفية سلوك جميع الأسعار عبر الزمن.

معوقات تركيب الأرقام القياسية

Problems in index Construction

تتمثل أهم العوائق في تركيب الأرقام القياسية في :

١- اختيار السلع التي يتضمنها الرقم القياسي.

٢- اختيار فتره الأساس.

٣- تحديد الطريقة الرياضية.

٤- تعيين الأوزان الواجب استخدامها.

السلع الواجب شمولها Items to Include

معظم مؤشرات الأسعار المرغوبة يتم نشرها بواسطة المؤسسات الحكومية أو القطاعات الخاصة الكبيرة المهتمة بذلك حيث يتم استخدامها في عدة طرق مختلفة. لذا فإنه ليس هناك هدف واحد لمؤشرات الأسعار. ولكن سنحاول الإجابة على الأسئلة الهادفة لأي مؤشر، والاستخدام الرئيسي للمؤشر يحدد السلع الواجب إدخالها في تركيب ذلك الرقم القياسي. فلو اعتبرنا مؤشر سعر المستهلك أو الرقم القياسي لسعر المستهلك والذي على سبيل المثال يعتبره كثير من المتنبئين مؤشر عام للرفاهية الاقتصادية للاقتصاد. ولكن هناك استعمالات أخرى له، فمستويات الأجور لمجموعات مختلفة من العاملين والمتقاعدين مرتبطة به، حيث تتم مراقبته من قبل الهيئات النقدية كمؤشر للتضخم وغير ذلك. وهذا الرقم القياسي له عدة قيود عند تطبيقه على مجموعة أهداف أخرى.

ومع ذلك فإن الغرض الرئيسي له هو تسجيل متوسط الحركة للأسعار الحالية على مر الزمن. وهذا هو الغرض الذي يشير إلى أي من سلاسل البيانات سوف تُشمل أو لا تُشمل في الرقم القياسي، وتحدد البيانات المشمولة في الرقم القياسي قيود هذا الرقم القياسي؛ إضافة على قوته. لذا فإننا نحتاج إلى أن نكون على يقين من فهم ماذا

يقيس الرقم القياسي ومن ثم عدم التعميم للحالات التي لا يستطيع الرقم القياسي التعامل معها.

وكيفية اختيار السلع لشمولها في الرقم القياسي يعتمد على أهمية تلك السلع. لذا فإننا لا نستطيع استخدام طريقه العينة العشوائية ؛ لأن كل سلعة أو خدمة لا يمكن اعتبارها كعينه ممثلة للأخرى. بدلاً من ذلك فإننا ندخل خصيصاً جميع السلع المهمة واستخدامها لتمثيل المجتمع.

وبعد تحديد السلع الواجب إدراجها في الرقم القياسي ربما نستخدم طرق معاينه متقدمة لتحديد السلع للمؤشر.

فترة الأساس The Base Period

الشهر أو السنة أو السلسلة من السنة والتي يقيس فيها الرقم القياسي التغير تدعى فترة الأساس. ونحن دائماً نضع رقم الرقم القياسي لفترة الأساس مساوياً لـ ١٠٠ ونعبر عن القيم للفترات الزمنية الأخرى كنسبة لفترة الأساس، ولذا فإننا نهدف لاختيار فتره طبيعية للأساس. وبالطبع قد نجد من الصعوبة تحديد فتره طبيعية لمعظم المتغيرات الاقتصادية، ولكن نحن لا نرغب في اختيار بعض الفترات التي تكون قمة أو قاع للسلسلة والتي تتقلب عبر الزمن. ولتركيب هذا الرقم القياسي فإنه ليس من الصعوبة عمله رياضياً، ولكنه يسبب تشوّهاً ؛ نظراً لأن استخدامه في المقارنة مبني على أوقات غير طبيعيه للبيانات إما القمم وإما القيعان.

معظم مؤشرات الحكومة الأمريكية تستخدم فترة من سنوات كأساس لأنها تعطي الأثر المتوسط للمتغيرات من سنة إلى أخرى. أيضاً نحن نرغب أن تكون فترة الأساس قريبة نوعاً ما من الحاضر، فمثلاً نفضل أساس من ١٩٩٠-١٩٩٤ على أساس من ١٩١٠-١٩١٤. وعندما تكون فترة الأساس تمثل وقت قديم جداً، فإننا قد

نسى ما هي الظروف الاقتصادية السائدة في ذلك الزمن وتكون المقارنة بهذه الفترة ليس لها معنوية ويصبح من الصعوبة تفسيرها نظراً للتغيرات الاجتماعية والتكنولوجية. لذا فإن معظم الهيئات الحكومية تغيّر فترات الأساس كل عقد أو قريب منه. أيضاً ، إذا كان ذلك ممكناً فإننا نرغب في اختيار فتره أساس معتادة والتي سيتم استخدامها في تركيب أرقام قياسية قريبة نسبياً ؛ وذلك لتسهيل عملية المقارنة.

أساليب التركيب الأساسية Basic Construction Techniques

معظم الأرقام القياسية المركبة عبارة عن أنواع تجميعية من حيث إن أكثر الأرقام القياسية البسيطة عبارة عن متوسطات نسبية. في الرقم القياسي البسيط ، الخطوة الأولى تشتمل على حساب النسبة للسلعة بقسمة قيمتها في السنة الحالية (ليست سنة الأساس) على قيمة سنة الأساس. وعند اشتمال الرقم القياسي على مجموعة سلع فإنه يتم أخذ متوسط النسب للحصول على الرقم القياسي العام وربما أيضاً تستخدم الأوزان كجزء من عملية إيجاد المتوسط. في ا الأرقام القياسية الحصريه أو التجميعية يتم جمع القيم لغير سنة الأساس ثم نقارنها بمجموع سنة الأساس وأيضاً ربما نستخدم الأوزان خلال عمليات التجميع.

الأوزان Weights

يتم استخدام أسلوب الأوزان ؛ لأن السلع المشمولة في الرقم القياسي عامة ليس لها نفس الدرجة من الأهمية. وفي بعض الأرقام القياسية مثل مؤشرات الأسعار تختلف الأهمية النسبية للسلع عبر الزمن ؛ نتيجة لتغير ظروف السوق. وهذا يسبب مشكلة عند تركيب الرقم القياسي للأشياء التي لا يوجد لها حلول مرضية. ونستخدم بصفة عامة الكميات كأوزان في الأرقام القياسية للأسعار مع الاهتمام بنوع الكميات هل هي لفترة الأساس أم للفترة الحالية أو تمثل متوسط للفترتين.

ومجموعة الكميات المختارة للترجيح تؤثر على قيمة الرقم القياسي. وكذلك في

حالة تأثير اختلاف شروط السوق على الطلب للسلع الحالية فإننا نهمل التغير إذا كانت فترة الأساس هي المستخدمة في الترجيح. أما إذا تم اختيار أوزان السنة الحالية فإننا نسبب بعض المشكلات باستخدام قيم الرقم القياسي للسنوات السابقة عندما لا يحدث أي تغير في الطلب، ... إلخ.

ومن الصعوبة اختيار الأوزان التي تعكس الجودة للسلع خاصة عندما يكون هناك تغيرات تكنولوجية. ومن الأمثلة على ذلك الرقم القياسي للسعر والذي يشتمل على الحاسبات الشخصية. فبينما الأسعار معتدلة نوعاً ما، فإن الحاسب الشخصي الأساسي الذي تم تسويقه بواسطة IBM في بدايات ١٩٨٠م كان بسعر مقارب لآخر موديل والذي يحتوى على معالج بسرعة أكبر ١٠٠ مرة مقارنة بمعالجات فترة الأساس، وكذلك الحال بالنسبة للمكونات الأخرى. لذا فإن الرقم القياسي للسعر الذي يحتوي على هذه السلعة يقلّل بدرجة كبيره من مستوى التغيرات السعرية.

تركيب الرقم القياسي Index Construction

بعد اختيار فترة الأساس للرقم القياسي، يتم اختيار نوع التركيب بسيط، نسبي، مرجح، تجميعي، ... إلخ . وإذا كان مرجح فما هو نوع الترجيح. ويعتمد الاختيار النهائي على نوع البيانات المتاحة لنا والتي يجب أن تكون بيانات زمنية دقيقة.

الأرقام القياسية البسيطة للأسعار Simple Price Indexes

نظراً لأن الأسعار تلعب أهمية كبيره في الأنشطة الاقتصادية، فإننا بصفه عامه مهتمين بتركيب الرقم القياسي للأسعار. ومعظم القياسيات البسيطة تصف التغيرات إما في السلع المفردة أو مجموعة من السلع. ويستخدم مؤشر الأسعار النسبي البسيط لشرح التغيرات في سلعة واحدة.

السعر النسبي البسيط Simple Price Relative

$$I_0 = [p_0/p_0]\times 100 = 100$$
$$I_n = [p_n/p_0]\times 100$$

الجزء الأول في المعادلة السابقة يوضح أن الرقم القياسي لسنة الأساس هو ١٠٠ ولكن الرقم القياسي لغير سنة الأساس، I_n عبارة عن نسبة مئوية تعبّر عن زيادة أو نقص في سعر السلعة من سعر سنة الأساس. فمثلاً نفترض أن سعر القمح المزرعي على المستوى الوطني يساوي ٣,٢٤ دولار للبوشل في عام ١٩٩٨م و ٣,٢٦ دولار في عام ١٩٩٩م و ٣,٤٥ دولار في ٢٠٠٠م. فإذا اخترنا السنة ١٩٩٨م لتكون هي سنه الأساس، فإنه يمكن تركيب السعر النسبي البسيط للثلاث السنوات كالتالي:

السعر النسبي لعام ١٩٩٨م : $[P_0 / P_0]\times 100 = [3.24 / 3.24]\times 100 = 100$

السعر النسبي لعام ١٩٩٩م : $[P_1 / P_0]\times 100 = [3.26 / 3.24]\times 100 = 100.6$

السعر النسبي لعام ٢٠٠٠م : $[P_2 / P_0]\times 100 = [3.45 / 3.24]\times 100 = 106.5$

لذا فإنه في عام ١٩٩٩م زاد سعر القمح المزرعي بنسبة ٠,٦٪ عن سعره في العام ١٩٩٨م ولكن في عام ٢٠٠٠م كانت الزيادة في السعر بنسبة ٦,٥ ٪ مقارنة بسعر القمح في عام ١٩٩٨م.

وعند الرغبة في حساب الرقم القياسي أو الرقم القياسي لأسعار الحبوب المزرعية على المستوى الوطني فإننا نستخدم الرقم القياسي التجميعي البسيط للسعر؛ وذلك للبيانات الواردة في الجدول رقم (١٣,١).

الجدول رقم (١٣.١). أسعار الحبوب المحلية وحسابات مؤشر السعر التجميعي.

الحبوب	الوحدات	سعر ١٩٩٨ P_0	سعر ٢٠٠٠ P_2	السعر النسبي P_2/P_0
القمح	بوشل	٣,٢٤ دولار	٣,٤٥ دولار	١,٠٦
الأرز	وزن مئوي	٥,٨٩ دولار	٦,٧٨ دولار	١,١٥
الذرة	بوشل	٢,٠٧ دولار	٢,٢٦ دولار	١,٠٩
الذرة الشامية	بوشل	١,٨٩دولار	٢,١٣ دولار	١,١٣
الشعير	بوشل	٢,٠٤ دولار	٢,٠٣ دولار	١,٠٠
الإجمالي	بوشل	١٥,١٣ دولار	١٦,٦٥ دولار	٥,٤٣

الرقم القياسي التجميعي البسيط للسعر *Simple Aggregate Price Index*

$$I_n = \left[\sum p_n / \sum p_0\right] \times 100$$

$$I_n = [16.65/15.13] \times 100 = 110$$

هذا الرقم القياسي الذي قيمته ١١٠ يشير إلى أن أسعار الحبوب المزرعية على المستوى الوطني زادت بنسبه ١٠٪ في عام ٢٠٠٠م مقارنة بسنة الأساس عام ١٩٩٨م. وهذا الرقم القياسي فشل في تحديد كيفية تغيّر تلك الأسعار ولم يعط أي اعتبار للأهمية النسبية للأنواع المختلفة للحبوب في الصناعة. وإذا تم تطبيق الأرقام القياسية النسبية فإنه يمكن التغلب على الاعتراض الأول.

المتوسط البسيط للأسعار النسبية *Simple Average of Price Relatives*

$$I_n = \sum(p_n/p_0)/N \times 100$$

وللبيانات في الجدول رقم (١٣,١) نحصل على:

$$I_2 = 5.43/5 \times 100 = 108.6$$

الأرقام القياسية المرجحة للأسعار Weighted Price indexes

في حالة الأرقام القياسية المرجحة، فإننا نحسب الأهمية النسبية لكل سلعة في الرقم القياسي.

لذا فإنه لمثال أسعار الحبوب المزرعية السابقة يمكن تركيب رقم قياسي يوضح الأهمية لكل نوع من الحبوب في الصناعة. وفي المتوسط المرجح للأرقام القياسية للأسعار النسبية فإننا نستخدم الإنفاق للترجيح؛ حيث إن الإنفاق يساوي السعر مضروباً في الكمية. ويمكن صياغة الرقم القياسي بطريقتين اعتماداً على استخدام الإنفاق في سنة الأساس أو السنة الحالية كترجيح. ويمكن حساب كلا الصيغتين اعتماداً على البيانات الواردة في الجدول رقم (١٣,٢).

المتوسط المرجح للأسعار النسبية باستخدام أوزان سنة الأساس

Weighted Average of Price Relatives With Base Year Weights

$$I_n = \frac{\sum \frac{p_n}{p_0}(p_0 q_0)}{\sum p_0 q_0} \times 100 \qquad (13.1)$$

لحساب العبارة الرياضية في البسط للرقم القياسي في المعادلة أعلاه نحتاج لضرب الأسعار النسبية الموضحة في العمود الأخير من الجدول رقم (١٣,١) بالإنفاق الموضح في العمود الأخير في الجدول رقم (١٣,٢) ثم نجمعها لنحصل على النتائج الموضحة في الجدول رقم (١٣,٣).

والحد الذي في المقام هو عبارة عن المجموع من أول إلى آخر العمود في الجدول رقم (١٣,٢).

الجدول رقم (١٣.٢). بيانات إنتاج وأسعار الحبوب المحلية وحسابات أوزان الإنفاق .

	الأسعار(دولار)			الكميات(مليون)		الإنفاق (دولار)	
الحبوب	**١٩٩٨**	**٢٠٠٠**	الوحدات	**١٩٩٨**	**٢٠٠٠**	**١٩٩٨**	**٢٠٠٠**
	P_0	P_2		q_o	q_2	p_oq_o	P_2q_2
القمح	٣,٢٤	٣,٤٥	بوشل	٢٤٦٧	٢٣٢١	٧٩٩٣,٠٨	٨٠٠٧,٤٥
الأرز	٥,٨٩	٦,٧٨	وزن مئوي	١٧٩,٧	١٩٧,٨	١٠٥٨,٤٣	١٣٤١,٠٨
الذرة	٢,٠٧	٢,٢٦	بوشل	٩٤٧٧	١٠١٠٣	١٩٦١٧,٣٩	٢٢٨٣٢,٧٨
الذرة الشامية	١,٨٩	٢,١٣	بوشل	٨٧٥	٦٤٩	١٦٥٣,٧٥	١٣٨٢,٣٧
الشعير	٢,٠٤	٢,٠٣	بوشل	٤٥٥	٣٧٥	٩٢٨,٢٠	٧٦١,٢٥
الإجمالي		١٦,٦٥	بوشل			٣١٢٥٠,٨٥	٣٤٣٢٤,٩٣

ومن ثم يكون الرقم القياسي كالتالي :

$$I_n = \frac{\sum \frac{p_n}{p_0}(p_0 q_0)}{\sum p_0 q_0} \times 100 = \frac{33{,}934.94}{31{,}250.85} \times 100 = 108.6$$

المتوسط المرجح للأسعار النسبية باستخدام أوزان السنة الحالية

weighted Average of Price Relatives With Current Year Weights

$$I_n = \frac{\sum \frac{p_n}{p_0}(p_n q_n)}{\sum p_n q_n} \times 100 \qquad \textbf{(13.2)}$$

هذا الرقم القياسي يستخدم الإنفاق في السنة الحالية كأوزان (المعادلة رقم 13.2) وبالنسبة لبيانات مثال الرقم القياسي المرجح لأسعار الحبوب المزرعية ضربنا الأسعار

النسبية في آخر عمود من الجدول رقم (١٢,١). بالأنفاق في السنة الحالية والموضح في العمود الأخير من الجدول رقم (١٣,٢) للحصول على الأسعار النسبية المرجحة والموضحة في العمود الأخير من الجدول رقم (١٣,٣). والمجموع للأسعار النسبية المرجحة هو البسط في الصيغة الرياضية للمؤشر بينما المقام عبارة عن مجموع الإنفاق في السنة الحالية والموضح في العمود الأخير من الجدول رقم (١٣,٢).

بالتعويض عن هذه الأرقام في المعادلة نحصل على :

$$I_n = \frac{\sum \frac{p_n}{p_0}(p_n q_n)}{\sum p_n q_n} \times 100 = \frac{37{,}314.15}{34{,}324.93} \times 100 = 108.7$$

ويلاحظ أن الرقم القياسي النسبي المرجح للذرة قريب نوعاً ما من السعر النسبي ؛ نظراً لأن الإنفاق على الذرة زاد في السنة الحالية مما أعطاه قوة تأثير على الرقم القياسي ككل.

الرقم القياسي المرجح للأسعار التجميعية باستخدام أوزان سنة الأساس

weighted Aggregative Price index With Base Year Weights

مؤشر لاسبير The Laspeyeres index

$$I_n = \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 \qquad (13.3)$$

مؤشر لاسبير (المعادلة رقم 13.3) يرجح كل سعر مباشرة بالكمية المستهلكة في سنة الأساس كما في الجدول رقم (١٣,٤). ولحساب مؤشر لاسبير فإننا نجمع العمود الأوسط في الجدول رقم (١٣,٤). ثم نقسمه على العمود الأخير من الجدول رقم (١٣,٢).

$$I_n = \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 = \frac{33,934.94}{31,250.85} \times 100 = 108.6$$

ونلاحظ أن هذا الرقم القياسي يعطي نفس الجواب المتحصل عليه باستخدام مؤشر الأسعار النسبية للوسط المرجح باستخدام سنة الأساس كترجيح ويجب تساوي ذلك ؛ نظراً لأن الرقم القياسيين لهما نفس الصياغة الرياضية.

الجدول رقم (١٣.٣). الأسعار النسبية المرجحة بأوزان سنة الأساس والسنة الحالية.

الحبوب	أوزان سنة الأساس $(p_2/p_o)(p_o q_o)$	أوزان السنة الحالية $(p_2/p_o)(p_2 q_2)$
القمح	٨٥١١,١٥	٨٥٢٦,٤٥
الأرز	١٢١٨,٣٧	١٥٤٣,٧٣
الذرة	٢١٤١٨,٠٢	٢٤٩٢٨,٥٤
الذرة الشامية	١٨٦٣,٧٥	١٥٥٧,٩١
الشعير	٩٢٣,٦٥	٧٥٧,٥٢
الإجمالي	٣٣٩٣٤,٩٤	٣٧٣١٤,١٥

الرقم القياسي المرجح للأسعار التجميعية باستخدام أوزان السنة الحالية
Weighted Aggregative Price index with Current Year Weights

مؤشر باتش The Paasche Index

$$I_n = \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100 \qquad (13.4)$$

عندما نأخذ إنفاق السنة الحالية ونقسمه على أسعار سنة الأساس مرجحة بالكميات للسنة الحالية نحصل على مؤشر باتش وتوضح المعادلة رقم (13.4) الصيغة الرياضية لمؤشر باتش للسعر.

لبيانات أسعار الحبوب المزرعيه السابقة تم قسمة إجمالي الإنفاق في السنة الحالية والموضح في العمود الأخير من الجدول رقم (١٣,٢) على البيانات الموضحة في العمود

الأخير من الجدول رقم (١٣،٤) لحساب الرقم القياسي كالتالي :

$$I_n = \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100 = \frac{34,324.93}{31,589.90} \times 100 = 108.7$$

الجدول رقم (١٣.٤).الأسعار التجميعية المرجحة بأوزان سنة الأساس والسنة الحالية.

الحبوب	$p_2 q_o$ (دولار)	$p_o q_2$ (دولار)
القمح	٨٥١١،١٥	٧٥٢٠،٠٤
الأرز	١٢١٨،٣٧	١١٦٥،٠٤
الذرة	٢١٤١٨،٠٢	٢٠٩١٣،٢١
الذرة الشامية	١٨٦٣،٧٥	١٢٢٦،٦١
الشعير	٩٢٣،٦٥	٧٦٥،٠٠
الإجمالي	٣٣٩٣٤،٩٤	٣١٥٨٩،٩

والرقم المتحصل عليه هو نفس الرقم الذي تم حسابه باستخدام المرجح المتوسط لمؤشر الأسعار النسبية باستخدام أوزان السنة الحالية ؛ نظراً لأن كلا الطريقتين لهما نفس الصيغة الرياضية.

وفي الحقيقة فإن هناك مؤشرات للكميات والإنفاق بالإضافة لمؤشرات الأسعار التي تم مناقشتها ولكن نفس أسلوب التركيب ينطبق عليهما والاختلاف فقط في البيانات.

تمارين Exercises

١- افترض أن القيم لأسعار السلة الغذائية السوقية للعام ١٩٩٩م و ٢٠٠٠م، كما هي في الجدول التالي :

السلعة	١٩٩٩م		٢٠٠٠م	
	السعر	الكمية	السعر	الكمية
الحليب	١،٢٥ دولار/ربع علبه	٢٥ ربع	١،٢٥ دولار/ربع علبه	٢٠
الخبز	١،١٥ دولار/خبزه	٣٠ خبزه	١،٢٥ دولار/خبزه	٣٥
البيض	٠،٩ دولار/درزن	١٠ درزن	٠،٨٠ دولار/درزن	١٥

أ) أنشئ الأسعار النسبية لكل سلعة.

ب) أوجد مؤشر السعر التجميعي البسيط.

ج) أوجد المتوسط البسيط لمؤشر الأسعار النسبية.

د) أوجد المتوسط المرجح لمؤشر الأسعار النسبية باستخدام أوزان سنة الأساس والسنة الحالية.

هـ) أوجد مؤشر سعر لاسبير وباتش.

٢- احسب المتوسط المرجح لمؤشر الأسعار النسبية باستخدام أوزان سنة أساس مرجحة لبيانات التفاح التالية :

السنة	السعر (دولار /بوشل)		الإنتاج (مليون بوشل)	
	فوجي	ماكنتوش	فوجي	ماكنتوش
١٩٩٨م	٢,٤٥	٢,١٥	١,٢٨	١,٤٠
١٩٩٩م	٢,٥٢	٢,٢٣	١,٣٢	١,٤٧
٢٠٠٠م	٢,٦٠	٢,٤١	١,٣١	١,٥٢

٣- إجمالي المبيعات لشركتك كانت ١٢ مليون دولار عام ١٩٩٨م و ١٨ مليون دولار عام ٢٠٠٠م. استخدمت الشركة الرقم القياسي التالي للانكماش السعري. كم نسبة زيادة المبيعات الإجمالية الحقيقية من عام ١٩٩٨م إلى ٢٠٠٠م.

السعر	الرقم القياسي
١٩٩٨	١٠٠
١٩٩٩	١٢٣
٢٠٠٠	١٥٠

٤- الجدول التالي يوضح مؤشر استخدام العمل المزرعي والذي تم إعداده على اعتبار أن عام١٩٨٩م هي سنة الأساس. أعد الرقم القياسي بحيث تكون سنة ١٩٩٩م = ١٠٠ هي سنة الأساس.

السنة	مؤشر استخدام العمل الزراعي
١٩٨٥	١١٠
١٩٨٦	١٠٦
١٩٨٧	١٠٢
١٩٨٨	١٠٢
١٩٨٩	١٠٠
١٩٩٠	٩٥
١٩٩١	٩٥
١٩٩٢	٩١
١٩٩٣	٨٥
١٩٩٤	٨٤
١٩٩٥	٨٦
١٩٩٦	٨٥
١٩٩٧	٨٥
١٩٩٨	٨٧
١٩٩٩	٨٢
٢٠٠٠	٨١

ملحق الجداول

الجدول رقم (١). الأرقام العشوائية.

APPENDIX TABLE 1 Random Numbers						
345769	953810	627280	423578	353511	899906	82708
549075	004410	059309	271243	403382	248735	972383
423480	950812	197145	556566	655917	046169	363201
554518	514280	950974	482196	058868	474936	724289
797165	670995	791954	188521	950156	086813	033365
062730	163375	602168	908350	360861	152201	966097
356756	519371	679389	371912	502903	936741	636775
700770	781547	916968	136999	801855	605975	295802
279584	733750	487151	116069	274869	416181	610911
862434	481154	391464	021094	761599	474456	582253
199585	167701	170778	934765	761328	275799	323046
048736	514507	977406	158840	846761	198016	933522
815218	609732	629295	517386	824505	676788	304971
643021	527212	492869	261844	914505	354436	355772
164332	245407	517804	422658	751712	583087	286872
174303	085157	308590	535846	503131	266915	465641
136325	414066	452293	649359	844625	674828	953396
117780	407444	426115	108970	621527	601599	652376
435697	245510	946158	934221	824917	509832	362638
912252	579474	848845	824321	049853	151126	052643
754438	658573	717914	040054	630638	264060	594641
322053	924909	048177	957012	801464	833319	978384
897199	125506	708669	408374	737887	906201	599469
046637	642050	435779	502427	027842	515775	811203
979346	721653	260190	842505	797017	157497	179041
202312	011976	373248	374293	802292	646914	171322
354014	356787	511271	904434	068589	329862	829316
682909	809290	793392	098004	120575	469925	112743
897690	572456	871574	465543	486529	507767	608677
139029	160636	417690	191242	625269	104858	020808
341447	723998	905614	519309	926345	240082	395043
415603	129727	894956	780924	227496	134056	023014
014881	496311	750082	707823	738906	157591	072396
827235	783798	324650	485324	568156	098331	768720
261607	730824	341940	259028	253973	145183	658110
527920	834376	972906	627959	654790	342497	593779

تابع الجدول رقم (١).

APPENDIX TABLE 1 Random Numbers (continued)						
835431	206253	467521	029822	700399	554652	450184
512651	743206	118787	587401	921517	015407	206860
376187	189133	154812	828785	667020	998697	579598
092530	869028	483691	165063	847894	041617	762976
238036	016856	290105	538530	079931	412195	838814
308168	717698	919814	092230	215657	469994	805803
773429	915639	900911	276895	149505	540379	224349
171626	601259	009905	572567	441960	299704	313987
180570	665625	424048	713009	830314	664642	521021
558715	965963	494210	875287	488595	898691	713010
345067	361180	989224	138905	355519	045847	746266
583819	310956	174728	099164	118461	758000	496302
615026	599459	722322	555090	572720	826686	456517
812358	389535	166779	441968	105639	632418	340890
784592	003651	279275	055646	341897	510689	026160
094619	636747	934082	787345	772825	603866	565688
450908	919891	157771	114333	710179	062848	615156
593546	728768	984323	290410	970562	906724	315005
873778	491131	209695	604075	783895	862911	762026
965705	317845	169619	921361	315606	990029	745251
311163	943589	540958	556212	760508	129963	236556
454554	284761	269019	924179	670780	389869	519229
124330	819763	596075	064570	495169	030185	866211
920765	122124	423205	596357	469969	072245	359269
183002	540547	312909	379818	464023	768381	377241
600135	865974	929756	162716	415598	878513	994633
235787	023117	895285	027055	943962	381112	530492
953379	655834	283102	836259	437761	391976	940853
009658	521970	537626	806052	715247	808585	252503
176570	849057	387097	311529	893745	450267	182626
747456	304530	931013	678688	270736	355032	400713
486876	631985	368395	154273	959983	672523	210456
987193	268135	867829	025419	301168	409545	131960
358155	950977	170562	246987	884126	785621	467942
021394	182615	049084	942153	278313	872709	693590
735047	428941	630704	893281	716045	267529	427605

الجدول رقم (٢). احتمالات ثنائي الحدين .

APPENDIX TABLE 2 Binominal Probabilities

n	*x*	*p*										
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
2	0	0.902	0.810	0.640	0.490	0.360	0.250	0.160	0.090	0.040	0.010	0.002
	1	0.095	0.180	0.320	0.420	0.480	0.500	0.480	0.420	0.320	0.180	0.095
	2	0.02	0.010	0.040	0.090	0.160	0.250	0.360	0.490	0.640	0.810	0.902
3	0	0.857	0.792	0.512	0.343	0.216	0.125	0.064	0.027	0.008	0.001	
	1	0.135	0.243	0.384	0.441	0.432	0.375	0.288	0.189	0.096	0.027	
	2	0.007	0.027	0.096	0.189	0.228	0.375	0.432	0.441	0.384	0.243	0.135
	3		0.001	0.008	0.027	0.064	0.125	0.216	0.343	0.512	0.729	0.857
4	0	0.815	0.656	0.410	0.240	0.130	0.062	0.026	0.008	0.002		
	1	0.171	0.292	0.410	0.412	0.346	0.250	0.154	0.076	0.026	0.004	
	2	0.014	0.049	0.154	0.265	0.346	0.375	0.346	0.265	0.154	0.049	0.014
	3		0.004	0.026	0.076	0.154	0.250	0.346	0.412	0.410	0.292	0.171
	4			0.002	0.008	0.026	0.062	0.130	0.240	0.410	0.656	0.815
5	0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002			
	1	0.204	0.328	0.410	0.360	0.259	0.156	0.077	0.028	0.006		
	2	0.021	0.073	0.205	0.309	0.346	0.312	0.230	0.132	0.051	0.008	0.001
	3	0.001	0.008	0.051	0.132	0.230	0.312	0.346	0.309	0.205	0.073	0.021
	4			0.006	0.028	0.077	0.156	0.259	0.360	0.410	0.328	0.204
	5				0.002	0.010	0.031	0.078	0.168	0.328	0.590	0.774
6	0	0.735	0.531	0.262	0.118	0.047	0.016	0.004	0.001			
	1	0.232	0.354	0.393	0.303	0.187	0.094	0.037	0.010	0.002		
	2	0.031	0.098	0.246	0.324	0.311	0.234	0.138	0.060	0.015	0.001	
	3	0.002	0.015	0.082	0.185	0.276	0.312	0.276	0.185	0.082	0.015	0.002
	4		0.001	0.015	0.060	0.138	0.234	0.311	0.324	0.246	0.098	0.031
	5			0.002	0.010	0.037	0.094	0.187	0.303	0.393	0.354	0.232
	6				0.001	0.004	0.016	0.047	0.118	0.262	0.531	0.735

تابع الجدول رقم (٢) .

APPENDIX TABLE 2 Binominal Probabilities (continued)

n	x	p										
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
7	0	0.698	0.478	0.210	0.082	0.028	0.008	0.002				
	1	0.257	0.372	0.367	0.247	0.131	0.055	0.017	0.004			
	2	0.041	0.124	0.275	0.318	0.261	0.161	0.077	0.025	0.004		
	3	0.004	0.023	0.115	0.227	0.290	0.273	0.194	0.097	0.029	0.003	
	4		0.003	0.029	0.097	0.194	0.373	0.290	0.227	0.115	0.023	0.004
	5			0.004	0.025	0.077	0.164	0.261	0.318	0.275	0.124	0.041
	6				0.004	0.017	0.055	0.131	0.247	0.367	0.372	0.257
	7					0.002	0.008	0.028	0.082	0.210	0.478	0.698
8	0	0.663	0.430	0.168	0.058	0.017	0.004	0.001				
	1	0.279	0.383	0.336	0.198	0.090	0.031	0.008	0.001			
	2	0.051	0.149	0.294	0.296	0.209	0.109	0.041	0.010	0.001		
	3	0.005	0.033	0.147	0.254	0.279	0.219	0.124	0.047	0.009		
	4		0.005	0.046	0.136	0.232	0.273	0.232	0.116	0.046	0.005	
	5			0.009	0.047	0.124	0.219	0.279	0.254	0.147	0.033	0.005
	6			0.001	0.010	0.041	0.109	0.209	0.296	0.294	0.149	0.051
	7				0.001	0.008	0.031	0.090	0.198	0.336	0.383	0.279
	8					0.001	0.004	0.017	0.058	0.168	0.430	0.663
9	0	0.630	0.387	0.134	0.040	0.010	0.002					
	1	0.299	0.387	0.302	0.156	0.060	0.018	0.004				
	2	0.063	0.172	0.302	0.267	0.161	0.070	0.021	0.004			
	3	0.008	0.045	0.176	0.267	0.251	0.164	0.074	0.021	0.003		
	4	0.001	0.007	0.066	0.172	0.251	0.246	0.167	0.074	0.017	0.001	
	5		0.001	0.017	0.074	0.167	0.246	0.251	0.172	0.066	0.007	0.001
	6			0.003	0.021	0.074	0.164	0.251	0.267	0.176	0.045	0.008
	7				0.004	0.021	0.070	0.161	0.267	0.302	0.172	0.063
	8					0.004	0.018	0.060	0.156	0.302	0.387	0.299
	9						0.002	0.010	0.040	0.134	0.387	0.630

تابع الجدول رقم (٢) .

APPENDIX TABLE 2 Binominal Probabilities (continued)

n	*x*	*p* 0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
10	0	0.599	0.349	0.107	0.028	0.006	0.001					
	1	0.315	0.387	0.268	0.121	0.040	0.010	0.002				
	2	0.075	0.194	0.302	0.33	0.121	0.044	0.011	0.001			
	3	0.010	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001		
	4	0.001	0.011	0.088	0.200	0.251	0.205	0.111	0.037	0.006		
	5		0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001	
	6			0.006	0.037	0.111	0.205	0.251	0.200	0.088	0.011	0.001
	7			0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057	0.010
	8				0.001	0.011	0.044	0.121	0.233	0.302	0.194	0.075
	9					0.002	0.010	0.040	0.121	0.268	0.387	0.315
	10						0.001	0.006	0.028	0.107	0.349	0.599
11	0	0.569	0.314	0.086	0.020	0.004						
	1	0.329	0.384	0.236	0.093	0.027	0.005	0.001				
	2	0.087	0.213	0.295	0.200	0.089	0.027	0.005	0.001			
	3	0.014	0.071	0.221	0.257	0.177	0.081	0.023	0.004			
	4	0.001	0.016	0.111	0.220	0.236	0.161	0.070	0.017	0.002		
	5		0.002	0.039	0.132	0.221	0.226	0.147	0.057	0.010		
	6			0.010	0.057	0.147	0.226	0.221	0.132	0.039	0.002	
	7			0.002	0.017	0.070	0.161	0.236	0.220	0.111	0.016	0.001
	8				0.004	0.023	0.081	0.177	0.257	0.221	0.071	0.014
	9				0.001	0.005	0.027	0.089	0.200	0.295	0.213	0.087
	10					0.001	0.005	0.027	0.093	0.236	0.384	0.329
	11							0.004	0.020	0.086	0.314	0.569

تابع الجدول رقم (٢) .

APPENDIX TABLE 2 Binominal Probabilities (continued)

n	*x*						*p*					
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
12	0	0.540	0.282	0.069	0.014	0.002						
	1	0.341	0.377	0.206	0.071	0.017	0.003					
	2	0.099	0.230	0.283	0.168	0.064	0.016	0.002				
	3	0.017	0.085	0.236	0.240	0.142	0.054	0.012	0.001			
	4	0.002	0.021	0.133	0.231	0.213	0.121	0.042	0.008	0.001		
	5		0.004	0.053	0.158	0.227	0.193	0.101	0.029	0.003		
	6			0.016	0.079	0.177	0.226	0.177	0.079	0.016		
	7			0.003	0.029	0.101	0.193	0.227	0.158	0.053	0.004	
	8			0.001	0.008	0.042	0.121	0.213	0.231	0.133	0.021	0.002
	9				0.001	0.012	0.054	0.142	0.240	0.236	0.085	0.017
	10					0.002	0.016	0.064	0.168	0.283	0.230	0.099
	11						0.003	0.017	0.071	0.206	0.377	0.341
	12							0.002	0.014	0.069	0.282	0.540
13	0	0.513	0.254	0.055	0.010	0.001						
	1	0.351	0.367	0.179	0.054	0.011	0.002					
	2	0.111	0.245	0.268	0.139	0.045	0.010	0.001				
	3	0.021	0.100	0.246	0.218	0.111	0.035	0.006	0.001			
	4	0.003	0.028	0.154	0.234	0.184	0.087	0.024	0.003			
	5		0.006	0.069	0.180	0.221	0.157	0.066	0.014	0.001		
	6		0.001	0.023	0.103	0.197	0.209	0.131	0.044	0.006		
	7			0.006	0.044	0.131	0.209	0.197	0.103	0.023	0.001	
	8			0.001	0.014	0.066	0.157	0.221	0.180	0.069	0.006	
	9				0.003	0.024	0.087	0.184	0.234	0.154	0.028	0.003
	10				0.001	0.006	0.035	0.111	0.218	0.246	0.100	0.021
	11					0.001	0.010	0.045	0.139	0.268	0.245	0.111
	12						0.002	0.011	0.054	0.179	0.367	0.351
	13							0.001	0.010	0.055	0.254	0.513

تابع الجدول رقم (٢) .

APPENDIX TABLE 2 Binominal Probabilities (continued)

n	*x*	*p*										
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
14	0	0.488	0.229	0.044	0.007	0.001						
	1	0.359	0.356	0.154	0.041	0.007	0.001					
	2	0.123	0.257	0.250	0.113	0.032	0.006	0.001				
	3	0.026	0.114	0.250	0.194	0.085	0.022	0.003				
	4	0.004	0.035	0.172	0.229	0.155	0.061	0.014	0.001			
	5		0.008	0.086	0.196	0.207	0.122	0.041	0.007			
	6		0.001	0.032	0.126	0.207	0.183	0.092	0.023	0.002		
	7			0.009	0.062	0.157	0.209	0.157	0.062	0.009		
	8			0.002	0.023	0.092	0.183	0.207	0.126	0.032	0.001	
	9				0.007	0.041	0.122	0.207	0.196	0.086	0.008	
	10				0.001	0.014	0.061	0.155	0.229	0.172	0.035	0.004
	11					0.003	0.022	0.085	0.194	0.250	0.114	0.026
	12					0.001	0.006	0.032	0.113	0.250	0.257	0.123
	13						0.001	0.007	0.041	0.154	0.356	0.359
	14							0.001	0.007	0.044	0.229	0.488
15	0	.463	0.206	0.035	0.005							
	1	0.366	0.343	0.132	0.031	0.005						
	2	0.135	0.267	0.231	0.092	0.022	0.003					
	3	0.031	0.129	0.250	0.170	0.063	0.014	0.002				
	4	0.005	0.043	0.188	0.219	0.127	0.042	0.007	0.001			
	5	0.001	0.010	0.103	0.206	0.186	0.092	0.024	0.003			
	6		0.002	0.043	0.147	0.207	0.153	0.061	0.012	0.001		
	7			0.014	0.081	0.177	0.196	0.118	0.035	0.003		
	8			0.003	0.035	0.118	0.196	0.177	0.081	0.014		
	9			0.001	0.012	0.061	0.153	0.207	0.147	0.043	0.002	
	10				0.003	0.024	0.092	0.186	0.206	0.103	0.010	0.001
	11				0.001	0.007	0.042	0.127	0.219	0.188	0.043	0.005
	12					0.002	0.014	0.063	0.170	0.250	0.129	0.031
	13						0.003	0.022	0.092	0.231	0.267	0.135
	14							0.005	0.031	0.132	0.343	0.366
	15								0.005	0.035	0.206	0.463

الجدول رقم (٣) . معاملات ثنائي الحدين .

APPENDIX TABLE 3 Binominal Coefficients, ${}_nC_r$

n	r								
	0	1	2	3	4	5	6	7	8
0	1								
1	1	1							
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
7	1	7	21	35	35	21	7	1	
8	1	8	28	56	70	56	28	8	1
9	1	9	36	84	126	126	84	36	9
10	1	10	45	120	210	252	210	120	45
11	1	11	55	165	330	462	462	330	165
12	1	12	66	220	495	792	924	792	495
13	1	13	78	286	715	1.287	1.716	1.716	1.287
14	1	14	91	364	1.001	2.002	3.003	3.432	3.003
15	1	15	105	455	1.365	3.003	5.005	6.435	6.435

For coefficients missing from the table, use the relation ${}_nC_r = {}_nC_{n-r}$

For example ${}_{12}C_{11} = {}_{12}C_1 = 12$

الجدول رقم (٤). قيم المضروب.

APPENDIX TABLE 4 Factorials	
n	n!
0	1
1	1
2	2
3	6
4	24
5	120
6	720
7	5.040
8	40.320
9	362.880
10	3.628.800
11	39.916.800
12	479.001.600
13	6.227.020.800
14	87.178.291.200
15	1.307.674.368.000

الجدول رقم (٥). قيم e^{-x}

APPENDIX TABLE 5 Values of e^{-x}							
x	e^{-x}	x	e^{-x}	x	e^{-x}	x	e^{-x}
0.0	1.000	2.5	0.0821	5.0	0.00674	7.5	0.000553
0.1	0.9048	2.6	0.0743	5.1	0.00610	7.6	0.000500
0.2	0.8187	2.7	0.0672	5.2	0.00552	7.7	0.000453
0.3	0.7408	2.8	0.0608	5.3	0.00499	7.8	0.000410
0.4	0.6703	2.9	0.0550	5.4	0.00452	7.9	0.000371
0.5	0.6065	3.0	0.0498	5.5	0.00409	8.0	0.000335
0.6	0.5488	3.1	0.0450	5.6	0.00370	8.1	0.000304
0.7	0.4966	3.2	0.0408	5.7	0.00335	8.2	0.000275
0.8	0.4493	3.3	0.0369	5.8	0.00303	8.3	0.000249
0.9	0.4066	3.4	0.0334	5.9	0.00274	8.4	0.000225
1.0	0.3679	3.5	0.0302	6.0	0.00248	8.5	0.000203
1.1	0.3329	3.6	0.0273	6.1	0.00224	8.6	0.000184
1.2	0.3012	3.7	0.0247	6.2	0.00203	8.7	0.000167
1.3	0.2725	3.8	0.0224	6.3	0.00184	8.8	0.000151
1.4	0.2466	3.9	0.0202	6.4	0.00166	8.9	0.000136
1.5	0.2231	4.0	0.0183	6.5	0.00150	9.0	0.000123
1.6	0.2019	4.1	0.0166	6.6	0.00136	9.1	0.000112
1.7	0.1827	4.2	0.0150	6.7	0.00123	9.2	0.000101
1.8	0.1653	4.3	0.0136	6.8	0.00111	9.3	0.000091
1.9	0.1496	4.4	0.0123	6.9	0.00101	9.4	0.000083
2.0	0.1353	4.5	0.0111	7.0	0.00091	9.5	0.000075
2.1	1.1224	4.6	0.0101	7.1	0.00082	9.6	0.000068
2.2	0.1108	4.7	0.0091	7.2	0.00075	9.7	0.000061
2.3	0.1003	4.8	0.0082	7.3	0.00068	9.8	0.000056
2.4	0.0907	4.9	0.0074	7.4	0.00061	9.9	0.000050

الجدول رقم (٦). توزيع بواسون لقيم مختارة للمتوسط.

APPENDIX TABLE 6 Poisson Distribution for Selected Value of mu

Values in the table are for the function:

$$P(X) = \frac{\mu^x e^{-\mu}}{x!}$$

P(X) for specified values of μ

μ

x	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0904	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0001	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0003	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

μ

x	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3662	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902
5	.0045	.0062	.0084	.0111	.0141	.0176	.0216	.0260	.0309	.0361
6	.0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	.0001	.0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	.0000	.0000	.0001	.0001	.0001	.0002	.0003	.0005	.0006	.0009
9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002

تابع – الجدول رقم (٦).

APPENDIX TABLE 6 Poisson Distribution for Selected Value of mu (continued).										
					μ					
x	**2.10**	**2.20**	**2.30**	**2.40**	**2.50**	**2.60**	**2.70**	**2.80**	**2.90**	**3.00**
0	.1125	.1108	.1103	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	..2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2337	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680
5	.0417	.0476	.0538	.0602	.0668	.0735	.0804	.0872	.0940	.1008
6	.0146	.0174	.0206	.0241	.0278	.0319	.0362	.0407	.0455	.0504
7	.0044	.0055	.0068	.0083	.0099	.0118	.0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	.0031	.0038	.0047	.0057	.0068	.0081
9	.0003	.0004	.0005	.0007	.0009	.0011	.0014	.0018	.0022	.0027
10	.0001	.0001	.0001	.0002	.0002	.0003	.0004	.0005	.0006	.0008
11	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002	.0002
12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
					μ					
x	**3.10**	**3.20**	**3.30**	**3.40**	**3.50**	**3.60**	**3.70**	**3.80**	**3.90**	**4.00**
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.0035	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.01771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1733	.1781	.1823	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.0278	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
					μ					
x	**4.10**	**4.20**	**4.30**	**4.40**	**4.50**	**4.60**	**4.70**	**4.80**	**4.90**	**5.00**
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0679	.0630	.0583	.0540	.0500	.0462	.0427	.0395	.0365	.0337
2	.1393	.1323	.1254	.1188	.1125	.1063	.1005	.0948	.0894	.0842

تابع – الجدول رقم (٦).

APPENDIX TABLE 6 Poisson Distribution for Selected Value of mu (continued).										
					μ					
x	**4.10**	**4.20**	**4.30**	**4.40**	**4.50**	**4.60**	**4.70**	**4.80**	**4.90**	**5.00**
3	.0904	.1852	.1798	.1743	.1687	.1631	.1574	.1517	.1460	.1404
4	.0951	.1944	.1933	.1917	.1898	.1875	.1849	.1820	.1789	.1755
5	.1600	.1633	.1662	.1687	.1708	.1725	.1738	.1747	.1753	.1755
6	.1093	.1143	.1191	.1237	.1281	.1323	.1362	.1398	.1432	.1462
7	.0640	.0686	.0732	.0778	.0824	.0869	.0914	.0959	.1002	.1044
8	.0328	.0360	.0393	.0248	.0463	.0500	.0537	.0575	.0614	.0653
9	.0150	.0168	.0188	.0209	.0232	.0255	.0281	.0307	.0334	.0363
10	.0061	.0071	.0081	.0092	.0104	.0118	.0132	.0147	.0164	.0181
11	.0023	.0027	.0032	.0037	.0043	.0049	.0056	.0064	.0073	.0082
12	.0008	.0009	.0011	.0013	.0016	.0019	.0022	.0026	.0030	.0034
13	.0002	.0003	.0004	.0005	.0006	.0007	.0008	.0009	.0011	.003
14	.0001	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005
15	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0002
					μ					
x	**5.10**	**5.20**	**5.30**	**5.40**	**5.50**	**5.60**	**5.70**	**5.80**	**5.90**	**6.00**
0	.0061	.0055	0.0050	.0045	.0041	.0037	.0033	.0030	.0027	.0025
1	.0311	.0287	.0265	.0244	.0225	.0207	.0191	.0176	.0162	.0149
2	.0793	.0746	.0701	.0659	.0618	.0580	.0544	.0509	.0477	.0446
3	.1348	.1293	.1239	.1185	.1133	.1082	.1033	.0985	.0938	.0892
4	.1719	.1681	.1641	.1600	.1558	.1515	.1472	.1428	.1383	.1339
5	.1753	.1748	.1740	.1728	.1714	.1697	.1678	.1656	.1632	.1606
6	.1490	.1515	.1537	.1555	.1571	.1584	.1594	.1601	.1605	.1606
7	.1086	.1125	.1163	.1200	.1234	.1267	.1298	.1326	.1353	.1377
8	.0692	.0731	.0771	.0810	.0849	.0887	.0925	.0962	.0998	.1033
9	.0392	.423	.0454	.0486	.0519	.0552	.0586	.0620	.0654	.0688
10	.0200	.0220	.0241	.0262	.0285	.0309	.0334	.0359	.0386	.0413
11	.0093	.0104	.0116	.0129	.0143	.0157	.0173	.0190	.0207	.0225
12	.0039	.0045	.0051	.0058	.0065	.0073	.0082	.0092	.0102	.0113
13	.0015	.0018	.0021	.0024	.0028	.0032	.0036	.0041	.0046	.0052
14	.0006	.0007	.0008	.0009	.0011	.0013	.0015	.0017	.0019	.0022
15	.0002	.0002	.0003	.0003	.0004	.0005	.0006	.0007	.0008	.0009
16	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003
17	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001

تابع – الجدول رقم (٦).

APPENDIX TABLE 6 Poisson Distribution for Selected Value of mu (continued).										
					μ					
x	**6.10**	**6.20**	**6.30**	**6.40**	**6.50**	**6.60**	**6.70**	**6.80**	**6.90**	**7.00**
0	.0022	.0020	.0018	.0017	.0015	.0014	.0012	.0011	.0010	.0009
1	.0137	.0126	.0116	.0106	.0098	.0090	.0082	.0076	.0070	.0064
2	.0417	.0390	.0364	.0340	.0318	.0296	.0276	.0258	.0240	.0223
3	.0848	.0806	.0765	.0726	.0688	.0652	.0617	.0584	.0552	.0521
4	.1294	.1249	.1205	.1161	.1118	.1076	.1034	.0992	.0952	.0912
5	.1579	.1549	.1519	.1487	.1454	.1420	.1385	.1349	.1314	.1277
6	.1605	.1601	.1595	.1586	.1575	.1562	.1546	.1529	.1511	.1490
7	.1399	.1418	.1435	.1450	.1462	.1472	.1480	.1486	.1489	.1490
8	.1066	.1099	.1130	.1160	.1188	.1215	.1240	.1263	.1284	.1304
9	.0723	.0757	.1791	.0825	.0858	.0891	.0923	.0954	.0985	.1014
10	.0441	.0469	.0498	.0528	.0558	.0588	.0618	.0649	.0679	.0710
11	.0244	.0265	.0285	.0307	.0330	.0353	.0377	.0401	.0426	.0452
12	.0124	.0137	.0150	.0164	.0179	.0194	.0210	.0227	.0245	.0263
13	.0058	.0065	.0073	.0081	.0089	.0099	.0108	.0119	.0130	.0142
14	.0025	.0029	.0033	.0037	.0041	.0046	.0052	.0058	.0064	.0071
15	.0010	.0012	.0014	.0016	.0018	.0020	.0023	.0026	.0029	.0033
16	.0004	.0005	.0005	.0006	.0007	.0008	.0010	.0011	.0013	.0014
17	.0001	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0006
18	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002
19	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001
					μ					
x	**7.10**	**7.20**	**7.30**	**7.40**	**7.50**	**7.60**	**7.70**	**7.80**	**7.90**	**8.00**
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0059	.0054	.0049	.0045	.0041	.0038	.0035	.0032	.0029	.0027
2	.0208	.0194	.0180	.0167	.0156	.0145	.0134	.0125	.0116	.0107
3	.0492	.0464	.0438	.0413	.0389	.0366	.0345	.0324	.0305	.0286
4	.0874	.0836	.0799	.0764	.0729	.0696	.0663	.0632	.0602	.0573
5	.1241	.1204	.1167	.1130	.1094	.1057	.1021	.0986	.0951	.0916
6	.1468	.1445	.1420	.1394	.1367	.1339	.1311	.1282	.1252	.1221
7	.1489	.1486	.1481	.1474	.1465	.1454	.1442	.1428	.1413	.1396
8	.1321	.1337	.1351	.1363	.1373	.1381	.1388	.1392	.1395	.1396
9	.1042	.1070	.1096	.1121	.1144	.1167	.1187	.1207	.1224	.1241
10	.0740	.0770	.0800	.0829	.0858	.0887	.0914	.0941	.0967	.0993
11	.0478	.0504	.0531	.0558	.0585	.0613	.0640	.0667	.0695	.0722

تابع – الجدول رقم (٦).

APPENDIX TABLE 6 Poisson Distribution for Selected Value of mu (continued).

					μ					
x	**7.10**	**7.20**	**7.30**	**7.40**	**7.50**	**7.60**	**7.70**	**7.80**	**7.90**	**8.00**
12	.0283	.0303	.0323	.0344	.0366	.0388	.0411	.0434	.0457	.0481
13	.0154	.0168	.0181	.0196	.0211	.0227	.0243	.0260	.0278	.0296
14	.0078	.0086	.0095	.0104	.0113	.0123	.0134	.0145	.0157	.0169
15	.0037	.0041	.0046	.0051	.0057	.0062	.0069	.0075	.0083	.0090
16	.0016	.0019	.0021	.0024	.0026	.0030	.0033	.0037	.0041	.0045
17	.0007	.0008	.0009	.0010	.0012	.0013	.0015	.0017	.0019	.0021
18	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
19	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003	.0003	.0004
20	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002
21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001
					μ					
x	**8.10**	**8.20**	**8.30**	**8.40**	**8.50**	**8.60**	**8.70**	**8.80**	**8.90**	**9.00**
0	.0003	.0003	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
1	.0025	.0023	.0021	.0019	.0017	.0016	.0014	.0013	.0012	.0011
2	.0100	.0092	.0086	.0079	.0074	.0068	.0063	.0058	.0054	.0050
3	.0269	.0252	.0237	.0222	.0208	.0195	.0183	.0171	.0160	.0150
4	.0544	.0517	.0491	.0466	.0443	.0420	.0398	.0377	.0357	.0337
5	.0882	.0849	.0816	.0784	.0752	.0722	.0692	.0663	.0635	.0607
6	.1192	.1160	.1128	.1097	.1066	.1034	.1003	.0972	.0941	.0911
7	.1378	.1358	.1338	.1317	.1294	.1271	.1247	.1222	.1197	.1171
8	.1395	.1392	.1388	.1382	.1375	.1366	.1356	.1344	.1332	.1318
9	.1256	.1269	.1280	.1290	.1299	.1306	.1311	.1315	.1317	.1318
10	.1017	.1040	.1063	.1084	.1104	.1123	.1140	.1157	.1172	.1186
11	.0749	.0776	.0802	.0928	.0853	.0878	.0902	.0925	.0948	.0970
12	.0505	.0530	.0555	.0579	.0604	.0629	.0654	.0679	.0703	.0728
13	.0315	.0334	.0354	.0374	.0395	.0416	.0438	.0459	.0481	.0504
14	.0182	.0196	.0210	.0225	.0240	.0256	.0272	.0289	.0306	.0324
15	.0098	.0107	.0116	.0126	.0136	.0147	.0158	.0169	.0182	.0194
16	.0050	.0055	.0060	.0066	.0072	.0079	.0086	.0093	.0101	.0109
17	.0024	.0026	.0029	.0033	.0036	.0040	.0044	.0048	.0053	.0058
18	.0011	.0012	.0014	.0015	.0017	.0019	.0021	.0024	.0026	.0029
19	.0005	.0005	.0006	.0007	.0008	.0009	.0010	.0011	.0012	.0014
20	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0005	.0006
21	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0002	.0003
22	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001

تابع – الجدول رقم (٦).

APPENDIX TABLE 6 Poisson Distribution for Selected Value of mu (continued).

	μ									
x	**9.10**	**9.20**	**9.30**	**9.40**	**9.50**	**9.60**	**9.70**	**9.80**	**9.90**	**10.00**
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	0005.
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	.0378
6	.0881	.0851	.0822	.0793	.0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	.1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263	.1251
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	.1752	.1776	.0799	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.1526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.1342	.0361	.0380	.0399	.0419	.0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	.0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	.0026	.0028	.0031	.0034	.0037
20	.0007	.0008	.0009	.0010	.0011	.0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	.0004	.0004
23	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

الجدول رقم (٧). التوزيع الطبيعي.

APPENDIX TABLE 7 Normal Curve Areas

مثال: نفترض أننا نرغب إيجاد الاحتمال لقيمة Z=1.57، ننظر تحت العمود الأول من اليسار حيث نختار القيمة 1.5 ثم نتحرك أفقياً حتى نصل للعمود 0.07 بحيث تكون نقطة التقاطع هي القيمة المطلوبة والتي تساوي 0.4418 والمكتوبة بالبنط العريض.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.2492	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	**.4418**	.4430	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	**.4750**	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817

تابع الجدول (٧).

APPENDIX TABLE 7 Normal Curve Areas (continued).										
Z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Also, for Z= 4.0, 5.0, and 6.0, the areas are 0.49997, 0.4999997, and 0.499999999.

الجدول رقم (٨). قيم t الحرجة

APPENDIX TABLE 8 Critical Values of t						
n	t_{100}	t_{050}	t_{025}	t_{010}	t_{005}	**v or d.f.**
2	3.078	6.314	12.706	31.821	63.657	1
3	1.886	2.920	4.303	6.965	9.925	2
4	1.638	2.353	3.182	4.541	5.841	3
5	1.533	2.132	2.776	3.747	4.604	4
6	1.476	2.015	2.571	3.365	4.032	5
7	1.440	1.943	2.447	3.143	3.707	6
8	1.415	1.895	2.365	2.998	3.499	7
9	1.397	1.860	2.306	2.896	3.355	8
10	1.383	1.833	2.262	2.821	3.250	9
11	1.372	1.812	2.228	2.764	3.169	10
12	1.363	1.796	2.201	2.718	3.106	11
13	1.356	1.782	2.179	2.681	3.055	12
14	1.350	1.771	2.160	2.650	3.012	13
15	1.345	1.761	2.145	2.624	2.977	14
16	1.341	1.753	2.131	2.602	2.947	15
17	1.337	1.746	2.120	2.583	2.921	16
18	1.333	1.740	2.110	2.567	2.898	17
19	1.330	1.734	2.101	2.552	2.878	18
20	1.328	1.729	2.093	2.539	2.861	19
21	1.325	1.725	2.086	2.528	2.845	20
22	1.323	1.721	2.080	2.518	2.831	21
23	1.321	1.717	2.074	2.508	2.819	22
24	1.319	1.714	2.069	2.500	2.807	23
25	1.318	1.711	2.064	2.492	2.797	24
26	1.316	1.708	2.060	2.485	2.787	25
27	1.315	1.706	2.056	2.749	2.779	26
28	1.314	1.703	2.052	2.473	2.771	27
29	1.313	1.701	2.048	2.467	2.763	28
30	1.311	1.699	2.045	2.462	2.756	29
Inf.	1.282	1.645	1.960	2.326	2.576	Inf.

الجدول رقم (٩). القيم الحرجة لمربع كاي

APPENDIX TABLE 9 Critical Values of Chi-Square

v or d.f.	$x^2 0.995$	$x^2 0.990$	$x^2 0.975$	$x^2 0.950$	$x^2 0.900$
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720
3	0.0717212	0.114832	0.215795	0.351846	0.584375
4	0.206990	0.297110	0.484419	0.710721	1.063623
5	0.411740	0.554300	0.831211	1.145476	1.61031
6	0.675727	0.872085	1.237347	1.63539	2.20413
7	0.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1979	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

تابع الجدول رقم (٩).

APPENDIX TABLE 9 Critical Values of Chi-Square (continued).					
$x^2 0.100$	$x^2 0.050$	$x^2 0.025$	$x^2 0.010$	$x^2 0.005$	v or d.f.
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9
15.9871	18.3070	20.4831	23.2093	25.1882	10
17.2750	19.6451	21.9200	24.7250	26.7569	11
18.5494	21.0261	23.3367	26.2170	28.2995	12
19.8119	22.3621	24.7356	27.6883	29.8194	13
21.0642	23.6848	26.1190	29.1413	31.3193	14
22.3072	24.9958	27.4884	30.5779	32.8013	15
23.5418	26.2962	28.8454	31.9999	34.2672	16
24.7690	27.5871	30.1910	33.4087	35.7185	17
25.9894	28.8693	31.5264	34.8053	37.1564	18
27.2036	30.1435	32.8523	36.1908	38.5822	19
28.4120	31.4104	34.1696	37.5662	39.9968	20
29.6151	32.6705	35.4789	38.9321	41.4010	21
30.8133	33.9244	36.7807	40.2894	42.7956	22
32.0069	35.1725	38.0757	41.6384	44.1813	23
33.1963	36.4151	39.3641	42.9798	45.5585	24
34.3816	37.6525	40.6465	44.3141	46.9278	25
35.5631	38.8852	41.9232	45.6417	48.2899	26
36.7412	40.1133	43.1944	46.9630	49.6449	27
37.9159	41.3372	44.4607	48.2782	50.9933	28
39.0875	42.5569	45.7222	49.5879	52.3356	29
40.2560	43.7729	46.9792	50.8922	53.6720	30
50.8050	55.7585	59.3417	63.6907	66.7659	40
63.1671	67.5048	71.4202	76.1639	79.4900	50
74.3970	79.0819	83.2976	88.3794	91.9517	60
85.5271	90.5312	95.0231	100.425	104.215	70
96.5782	101.879	106.629	112.329	116.321	80
107.565	113.145	118.136	124.116	128.299	90
118.498	124.342	129.561	135.807	140.169	100

الجدول رقم (١٠) . قيم توزيع F

الخط العادي لمستوى معنوية ٥٪ والخط العريض لمستوى معنوية ١٪

APPENDIX TABLE 10 5 and 1 Percent Values of the *F* Distribution

5% (Roman type) and 1% (boldface type)

	Degrees of Freedom (for greater mean square)																								
d.f.	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	d.f.
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254	1
	4.052	**4.999**	**5.403**	**5.625**	**5.764**	**5.859**	**5.928**	**5.981**	**6.022**	**6.056**	**6.082**	**6.106**	**6.142**	**6.169**	**6.208**	**6.234**	**6.261**	**6.286**	**6.302**	**6.323**	**6.334**	**6.352**	**6.361**	**6.366**	
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50	2
	98.49	**99.00**	**99.17**	**99.25**	**99.30**	**99.33**	**99.36**	**99.37**	**99.39**	**99.40**	**99.41**	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.48**	**99.48**	**99.49**	**99.49**	**99.49**	**99.49**	**99.50**	
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53	3
	34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**	**27.13**	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.35**	**26.27**	**26.23**	**26.18**	**26.14**	**26.12**	
4	7.71	6.24	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63	4
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.54**	**14.45**	**14.37**	**14.24**	**14.15**	**14.02**	**13.93**	**13.83**	**13.74**	**13.69**	**13.61**	**13.57**	**13.52**	**13.48**	**13.46**	
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36	5
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.29**	**10.15**	**10.05**	**9.96**	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.07**	**9.04**	**9.02**	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67	6
	13.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**	**7.79**	**7.72**	**7.60**	**7.52**	**7.39**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.94**	**6.90**	**6.88**	
7	5.59	4.74	4.34	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23	7
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.71**	**6.62**	**6.54**	**6.47**	**6.35**	**6.27**	**6.15**	**6.07**	**5.98**	**5.90**	**5.85**	**5.78**	**5.75**	**5.70**	**5.67**	**5.65**	
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93	8
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.91**	**5.82**	**5.74**	**5.67**	**5.56**	**5.48**	**5.36**	**5.28**	**5.20**	**5.11**	**5.06**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**	
9	512	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71	9
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.35**	**5.26**	**5.18**	**5.11**	**5.00**	**4.92**	**4.80**	**4.73**	**4.64**	**4.56**	**4.51**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54	10
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.21**	**5.06**	**4.95**	**4.85**	**4.78**	**4.71**	**4.60**	**4.52**	**4.41**	**4.33**	**4.25**	**4.17**	**4.12**	**4.05**	**4.01**	**3.96**	**3.93**	**3.91**	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40	11
	9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**	**4.88**	**4.74**	**4.63**	**4.54**	**4.46**	**4.40**	**4.29**	**4.21**	**4.10**	**4.02**	**3.94**	**3.86**	**3.80**	**3.74**	**3.70**	**3.65**	**3.62**	**3.60**	
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.16	2.42	2.40	2.36	2.35	2.32	2.31	2.30	12
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.65**	**4.50**	**4.39**	**4.30**	**4.22**	**4.16**	**4.05**	**3.98**	**3.86**	**3.78**	**3.70**	**3.61**	**3.56**	**3.49**	**3.46**	**3.41**	**3.38**	**3.36**	

تابع الجدول رقم (١٠) .

APPENDIX TABLE 10 5 and 1 Percent Values of the *F* Distribution (continued).

5% (Roman type) and 1% (boldface type)

	Degrees of Freedom (for greater mean square)																							
d.f.	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	d.f.
13	4.67	3.80	3.41	3.18	3.02	2.92	2.0	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21	13
	9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**	**4.02**	**3.96**	**3.85**	**3.78**	**3.67**	**3.59**	**3.51**	**3.42**	**3.37**	**3.30**	**3.27**	**3.21**	**3.18**	**3.16**	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13	14
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**	**3.86**	**3.80**	**3.70**	**3.62**	**3.51**	**3.43**	**3.34**	**3.26**	**3.21**	**3.14**	**3.11**	**3.06**	**3.02**	**3.00**	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07	15
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**	**3.73**	**3.67**	**3.56**	**3.48**	**3.26**	**3.29**	**3.20**	**3.12**	**3.07**	**3.00**	**2.97**	**2.92**	**2.89**	**2.87**	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01	16
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.78**	**3.69**	**3.61**	**3.55**	**3.45**	**3.37**	**3.25**	**3.18**	**3.10**	**3.01**	**2.96**	**2.98**	**2.86**	**2.80**	**2.77**	**2.75**	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	17
	8.40	**6.11**	**5.18**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.68**	**3.59**	**3.52**	**3.45**	**3.35**	**3.27**	**3.16**	**3.08**	**3.00**	**2.92**	**2.86**	**2.79**	**2.76**	**2.70**	**2.67**	**2.65**	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.93	1.92	18
	8.28	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.85**	**3.71**	**3.60**	**3.51**	**3.44**	**3.37**	**3.27**	**3.19**	**3.07**	**3.00**	**2.91**	**2.83**	**2.78**	**2.71**	**2.68**	**2.62**	**2.59**	**2.57**	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88	19
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**5.52**	**3.43**	**3.36**	**3.30**	**3.19**	**3.12**	**3.00**	**2.92**	**2.84**	**2.76**	**2.70**	**2.63**	**2.60**	**2.54**	**2.51**	**2.49**	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	20
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.71**	**3.56**	**3.45**	**3.37**	**3.30**	**3.23**	**3.13**	**3.05**	**2.94**	**2.86**	**2.77**	**2.69**	**2.63**	**2.56**	**2.53**	**2.47**	**2.44**	**2.42**	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	21
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.65**	**3.51**	**3.40**	**3.31**	**3.24**	**3.17**	**3.07**	**2.99**	**2.88**	**2.80**	**2.72**	**2.63**	**2.58**	**2.51**	**2.47**	**2.42**	**2.38**	**2.36**	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	22
	7.94	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.35**	**3.26**	**3.18**	**3.12**	**3.02**	**2.94**	**2.83**	**2.75**	**2.67**	**2.58**	**2.53**	**2.46**	**2.42**	**2.37**	**2.33**	**2.31**	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	23
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	**3.71**	**3.54**	**3.41**	**3.30**	**3.21**	**3.14**	**3.07**	**2.97**	**2.89**	**2.78**	**2.70**	**2.62**	**2.53**	**2.48**	**2.41**	**2.37**	**2.32**	**2.28**	**2.26**	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	24
	7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.50**	**3.36**	**3.25**	**3.17**	**3.09**	**3.03**	**2.93**	**2.85**	**2.74**	**2.66**	**2.58**	**2.49**	**2.44**	**2.36**	**2.33**	**2.27**	**2.23**	**2.21**	

تابع الجدول رقم (١٠) .

APPENDIX TABLE 10 5 and 1 Percent Values of the *F* Distribution (continued).

5% (Roman type) and 1% (boldface type)

													Degrees of Freedom (for greater mean square)												
d.f.	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	d.f.
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	25
	7.77	**5.57**	**4.68**	**4.18**	**3.86**	**3.63**	**3.46**	**3.32**	**3.21**	**3.13**	**3.05**	**2.99**	**2.89**	**2.81**	**2.70**	**2.62**	**2.54**	**2.45**	**2.40**	**2.32**	**2.29**	**2.23**	**2.19**	**2.17**	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	26
	7.725	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.17**	**3.09**	**3.02**	**2.96**	**2.86**	**2.77**	**2.66**	**2.58**	**2.50**	**2.41**	**2.36**	**2.28**	**2.25**	**2.19**	**2.15**	**2.13**	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67	27
	7.68	**5.49**	**4.60**	**4.11**	**3.79**	**3.56**	**3.39**	**3.26**	**3.14**	**3.06**	**2.98**	**2.93**	**2.83**	**2.74**	**2.63**	**2.55**	**2.47**	**2.38**	**2.33**	**2.25**	**2.21**	**2.16**	**2.12**	**2.10**	
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65	28
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.11**	**3.03**	**2.95**	**2.90**	**2.80**	**2.71**	**2.60**	**2.52**	**2.44**	**2.35**	**2.30**	**2.22**	**2.18**	**2.13**	**2.09**	**2.06**	
29	4.18	3.33	2.93	2.70	2.54	23.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64	29
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.08**	**3.00**	**2.92**	**2.87**	**2.77**	**2.68**	**2.57**	**2.49**	**2.41**	**2.32**	**2.27**	**2.19**	**2.15**	**2.10**	**2.06**	**2.03**	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62	30
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.06**	**2.98**	**2.90**	**2.84**	**2.74**	**2.66**	**2.55**	**2.47**	**2.38**	**2.29**	**2.24**	**2.16**	**2.13**	**2.07**	**2.03**	**2.01**	
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59	32
	7.50	**5.34**	**4.46**	**3.97**	**3.66**	**3.42**	**3.25**	**3.12**	**3.01**	**2.94**	**2.86**	**2.80**	**2.70**	**2.62**	**2.51**	**2.42**	**2.34**	**2.25**	**2.20**	**2.12**	**2.08**	**2.02**	**1.98**	**1.96**	
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57	34
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**6.38**	**3.21**	**3.08**	**2.97**	**2.89**	**2.82**	**2.76**	**2.66**	**2.58**	**2.47**	**2.38**	**2.30**	**2.21**	**2.15**	**2.08**	**2.04**	**1.98**	**1.94**	**1.91**	
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55	36
	7.39	**5.25**	**4.38**	**3.89**	**3.58**	**3.35**	**3.18**	**3.04**	**2.94**	**2.86**	**2.78**	**2.72**	**2.62**	**2.54**	**2.43**	**2.35**	**2.26**	**2.17**	**2.12**	**2.04**	**2.00**	**1.94**	**1.90**	**1.87**	
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53	38
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.91**	**2.82**	**2.75**	**2.69**	**2.59**	**2.51**	**2.40**	**2.32**	**2.22**	**2.14**	**2.08**	**2.00**	**1.97**	**1.90**	**1.86**	**1.84**	
40	4.07	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51	40
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.88**	**2.80**	**2.73**	**2.66**	**2.56**	**2.49**	**2.37**	**2.29**	**2.20**	**2.11**	**2.05**	**1.97**	**1.94**	**1.88**	**1.84**	**1.81**	
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49	42
	7.27	**5.15**	**4.29**	**3.80**	**3.49**	**3.26**	**3.10**	**2.96**	**2.86**	**2.77**	**2.70**	**2.64**	**2.54**	**2.46**	**2.35**	**2.26**	**2.17**	**2.08**	**2.02**	**1.94**	**1.91**	**1.85**	**1.80**	**1.78**	

تابع الجدول رقم (١٠) .

APPENDIX TABLE 10 5 and 1 Percent Values of the *F* Distribution (continued).

5% (Roman type) and 1% (boldface type)

	Degrees of Freedom (for greater mean square)																								
d.f.	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	d.f.
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48	44
	7.24	**5.12**	**4.26**	**3.78**	**3.46**	**3.24**	**3.07**	**2.94**	**2.84**	**2.75**	**2.68**	**2.62**	**2.52**	**2.44**	**2.32**	**2.24**	**2.15**	**2.06**	**2.00**	**1.92**	**1.88**	**1.82**	**1.78**	**1.75**	
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46	46
	7.21	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.82**	**2.73**	**2.66**	**2.60**	**2.50**	**2.42**	**2.30**	**2.22**	**2.13**	**2.04**	**1.98**	**1.90**	**1.86**	**1.80**	**1.76**	**1.72**	
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45	48
	7.19	**5.08**	**4.22**	**3.74**	**3.42**	**3.20**	**3.04**	**2.90**	**2.80**	**2.71**	**2.64**	**2.58**	**2.48**	**2.40**	**2.28**	**2.20**	**2.11**	**2.02**	**1.96**	**1.88**	**1.84**	**1.78**	**1.73**	**1.70**	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44	50
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.78**	**2.70**	**2.62**	**2.56**	**2.46**	**2.39**	**2.26**	**2.18**	**2.10**	**2.00**	**1.94**	**1.86**	**1.82**	**1.76**	**1.71**	**1.68**	
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41	55
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**	**2.59**	**2.53**	**2.43**	**2.35**	**2.23**	**2.15**	**2.06**	**1.96**	**1.90**	**1.82**	**1.78**	**1.71**	**1.66**	**1.64**	
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39	60
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**	**2.56**	**2.50**	**2.40**	**2.32**	**2.20**	**2.12**	**2.03**	**1.93**	**1.87**	**1.79**	**1.74**	**1.68**	**1.63**	**1.60**	
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37	65
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.79**	**2.70**	**2.61**	**2.54**	**2.47**	**2.37**	**2.30**	**2.18**	**2.09**	**2.00**	**1.90**	**1.84**	**1.76**	**1.71**	**1.64**	**1.60**	**1.56**	
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35	70
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.67**	**2.59**	**2.51**	**2.45**	**2.35**	**2.28**	**2.15**	**2.07**	**1.98**	**1.88**	**1.82**	**1.74**	**1.69**	**1.62**	**1.56**	**1.53**	
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32	80
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**	**2.48**	**2.41**	**2.32**	**2.24**	**2.11**	**2.03**	**1.94**	**1.84**	**1.78**	**1.70**	**1.65**	**1.57**	**1.52**	**1.49**	
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28	100
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.59**	**2.51**	**2.43**	**2.36**	**2.26**	**2.19**	**2.06**	**1.98**	**1.89**	**1.79**	**1.73**	**1.64**	**1.59**	**1.51**	**1.46**	**1.43**	
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25	125
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.56**	**2.47**	**2.40**	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.75**	**1.68**	**1.59**	**1.54**	**1.46**	**1.40**	**1.37**	

تابع الجدول رقم (١٠).

APPENDIX TABLE 10 5 and 1 Percent Values of the *F* Distribution (continued).

5% (Roman type) and 1% (boldface type)

	Degrees of Freedom (for greater mean square)																								
d.f.	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	d.f.
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22	150
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2.76**	**2.62**	**2.53**	**2.44**	**2.37**	**2.30**	**2.20**	**2.12**	**2.00**	**1.91**	**1.83**	**1.72**	**1.66**	**1.56**	**1.51**	**1.43**	**1.37**	**1.33**	
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19	200
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.50**	**2.41**	**2.34**	**2.28**	**2.17**	**2.09**	**1.97**	**1.88**	**1.79**	**1.69**	**1.62**	**1.53**	**1.48**	**1.39**	**1.33**	**1.28**	
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13	400
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.46**	**2.37**	**2.29**	**2.23**	**2.12**	**2.04**	**1.92**	**1.84**	**1.74**	**1.64**	**1.57**	**1.47**	**1.42**	**1.32**	**1.24**	**1.19**	
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08	1000
	6.66	**4.62**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**	**2.26**	**2.20**	**2.09**	**2.01**	**1.89**	**1.81**	**1.71**	**1.61**	**1.54**	**1.44**	**1.38**	**1.28**	**1.19**	**1.11**	
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00	∞
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**	**2.24**	**2.18**	**2.07**	**1.99**	**1.87**	**1.79**	**1.69**	**1.59**	**1.52**	**1.41**	**1.36**	**1.25**	**1.15**	**1.00**	

The function $F = e$, with exponent $2z$, is computed in part from Fisher's table VI (7). Additional entries are by interpolation, mostly graphical.

الجدول رقم (١١). مربعات لاتينية مختارة.

APPENDIX TABLE 11 Selected Latin Squares

3 x 3	4 x 4			
	1	2	3	4
A B C	A B C D	A B C D	A B C D	A B C D
C A B	D A B C	B A D C	B D A C	B A D C
B C A	C D A B	D C B A	C A D B	C D A B
	B C D A	C D B A	D C B A	D C B A

5 x 5	6 x 6	7 x 7
A B C D E	A B C D E F	A S C D E F G
E A B C D	F A B C D E	G A B C D E F
C D E A B	E F A B C D	F G A B C D E
D E A B C	D E F A B C	E F G A B C D
	C D E F A B	D E F G A B C
	B C D E F A	C D E F G A B
		B C D E F G A

8 x 8	9 x 9
A B C D E F G H	A B C D E F G H I
H A B C D E F G	I A B C D E F G H
G H A B C D E F	H I A B C D E F G
F G H A B C D E	G H I A B C D E F
E F G H A B C D	F G H I A B C D E
D E F G H A B C	E F G H I A B C D
C D E F G H A B	D E F G H I A B C
B C D E F G H A	C D E F G H I A B
	B C D E F G H I A

الجدول رقم (١٢). القيم الحرجة لـــ **r** في اختبار التتابع.

APPENDIX TABLE 12 Critical Values of *r* in the Runs Test

The body of the table contains critical values of *r* at the 0.05 level of significance for values of n_1 and n_2

$n_1 \mid n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	8	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

الجدول رقم (١٣). القيم الحرجة لمعاملات ارتباط الرتب لسبيرمان.

APPENDIX TABLE 13 Critical Values of Spearman's Rank Correlation Coefficient

n	α = 0.10	α = 0.05	α – 0.02	α = 0.01
5	0.900			
6	0.829	0.886	0.943	
7	0.714	0.786	0.893	
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.277	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478

مسرد المصطلحات

أولاً: عربي – انجليزي

أ

الاتجاه الزمني Secular trend : حركة انسيابية تصاعدية أو تنازلية تميز السلسلة الزمنية خلال فترة طويلة من الزمن، مثل ١٠ سنوات أو أكثر.

الاتحاد Union : العناصر الواردة في مجموعة واحدة أو مجموعة ثانية أو تلك المشتركة في كلتا المجموعتين.

الاحتمال Probability : تعبر عن نسبة عدد الحوادث المفضلة لحدث A إلى عدد الحوادث في التجربة.

الاحتمالات الشرطية Conditional probabilities : الاحتمالات المصاحبة للأحداث في المجتمعات الفرعية.

الاحتمالات القبلية A priori Probabilities: هي الاحتمالات التي تحدد باستخدام

النظرية أو التقدير الحدسي وهي عكس التجريبي.

الأحداث متساوية الاحتمالات Equiprobable events : إذا لم يكن هناك سبباً لتفضيل نتيجة معينة للتجربة ، فإنه يجب اعتبار أن جميع النتائج متساوية الاحتمالات.

الإحصاء الوصفي Descriptive statistics : بيانات رقمية.

الإحصاءات Statistics : متغيرات العينة.

الاختبارات اللامعلمية Nonparametric tests : الاختبارات التي لا يكون هناك رغبة في وضع فروض حول شكل توزيع المعاينة للإحصاءات محل الاختبار.

الاختلاف المفسر Explained Variation : النسبة في الاختلاف الكلي في المتغير العشوائي التي يتم تفسيرها بمجموعة من المتغيرات المستقلة Xs .

الاختلاف داخل العينة Variation within samples : مصطلح يستخدم عادة في تحليل التباين باتجاه واحد للإشارة إلى مجموع مربعات الخطاء أو متوسط مربعات الخطاء.

الارتباط الخطي Multicollinearity : يحدث عندما تكون المتغيرات المستقلة في النموذج الخطي مرتبطة داخلياً بدرجة كبيرة.

ارتباط الرتب Rank correlation : يحسب الارتباط العادي باستخدام الرتب للبيانات بدلا من المشاهدات الأصلية.

الارتباط Correlation : مقياس لدرجة العلاقة بين متغيرين عشوائيين ، x و y. وتكون مرتبطة بناء على حركتها معاً إما في اتجاه واحد أو اتجاه متعاكس.

الاستدلال الإحصائي Statistical inference : تطوير وتطبيق الأساليب والتقنيات المتعلقة بجمع وتلخيص وتفسير البيانات العددية لذا فإن الاعتماد على الخلاصة المرسومة من هذه الإجراءات يمكن تقييمه بصورة موضوعية باستخدام العبارات الاحتمالية.

الالتواء Skewness : التماثل النسبي للتوزيع. إذا كان التوزيع متماثل الشكل تماماً فليس هناك التواء. إما إذا كان للتوزيع طرف أيمن فإنه ذو التواء موجب أو ملتو لليمين، وإذا كان له طرف في اليسار فإنه ذو التواء سالب أو ملتو لليسار.

الانحدار الخطي المتعدد Multiple linear regression : معادلة تشرح الاختلافات في y باستخدام مجموعة من المتغيرات المستقلة x (أكثر من واحد).

الانحراف الربيعي Quartile deviation : مقياس للتشتت يستخدم فقط مع المتوسط ؛ ويشير إلى تشتت البيانات في منتصف الوسط للتوزيع.

الانحـــراف المعيـــاري Standard deviation : مقياس للتشتت يستخدم مع المتوسط الحسابي. ويتم حساب قيمته بناء على جميع المشاهدات في مجموعة البيانات. ويحسب بأخذ الجذر التربيعي للمتوسط الحسابي لمربعات انحراف القيم عن متوسطها.

ب

البـــواقي Residual : الفرق بين القيمة المشاهدة لـ Y والقيمة المقدرة لـ Y من خط الانحدار ، ويرمز لها بالرمز e_i.

بيانات اسمية Nominal data : البيانات تكون أسمية فقط. فمثلا يتم تصنيف بيانات الأبقار حسب أسمائها ، عجول ، تبيعه ، تبيع ، أبقار ، ثيران ، ... إلخ.

البيانات الترتيبيـــة Ordinal data : بيانات تعبر عن علاقة الترتيب وبالتالي يمكن استخدام الرتب. فمثلاً المستهلك ربما يرتب تفضيله للطعام ، ...إلخ.

البيانات القطاعية Cross section : مجموعة من البيانات الإحصائية التي تم جمعها خلال فترة زمنية محددة من مساحة واسعة.

البيانات المزدوجة Paired data : تكون البيانات غير المستقلة نظراً لوجود بعض العوامل التي تجعل بينها علاقة ؛ فمثلاً البيانات المجموعة لمحاولات متنوعة لمزارع مختلفة عبارة عن بيانات غير مستقلة نظراً لأن المهارات الإدارية للمزارعين والظروف الحقلية...إلخ ، مختلفة من مزرعة إلى مزرعة وبذلك تؤثر على اختلاف الإنتاجية.

ت

التباين Variance : مربع الانحراف المعياري ، وأيضاً يسمى متوسط المربعات في تحليل التباين. ويجب أن نستخدم التباين في الحسابات الرياضية ، ثم بعد ذلك نحسب الانحراف المعياري إذا كان هو المطلوب.

التبديل Permutation : أي ترتيب تسلسلي لمجموعة أو مجموعة أشياء.

التجربة Experiment : أي عملية من الملاحظات أو الحصول على بيانات.

التحيز Bias : يمكن تطبيقه على أي مقدر ، وتكون قيمته مساوية للصفر إذا كانت القيمة المتوقعة للمقدر مساوية لمقدر المجتمع. مثلا ، إذا كان $E(\overline{x}) = \mu$ لمتوسط العينة ، وما عدا ذلك فإن قيمة التحيز تساوي القيمة المتوقعة مطروحاً منها المقدّرة.

التشتت Dispersion : مقياس لتمثيل أو صحة المتوسط ، ويعبر عن القيمة التي توضح مدى انتشار أو تركز القيم حول متوسطها.وكلما كان التشتت عن المتوسط صغير كلما كان المتوسط أكثر تمثيلاً وعليه نستطيع استخدام المتوسط كرقم لعرض كامل التوزيع.

التغيرات الدورية Cyclical fluctuations : حركة الدورات الاقتصادية التي تعكس توسع الاقتصاد من فترة الكساد المنخفضة من خلال الازدهار الاقتصادي ثم بعد ذلك تنكمش باتجاه فترة كساد أخرى.

الـــتغيرات الموسميـــة Seasonal variation : دورات قصيرة الأجل تحدث للبيانات وتستكمل دورتها في غضون السنة، وتتكرر سنوياً.

التقاطع Intersection : المنطقة المشتركة لمجموعتين أو أكثر، أو العناصر المشتركة لهذه المجموعات

تقدير النقطة Point estimate : قيمة واحدة للمقدر؛ مثلاً، $\bar{x} = 2$.

التقديرات Estimates : قيم معينة للمتغيرات العشوائية.

التكرار المتجمع Ogive : منحنى التكرار التجميعي ويكون على شكل حرف S حيث توضع التكرارات المتجمعة على المحور الرأسي y وتوضع الحدود العليا للفئات على المحور الأفقي x. ويمكن تجميع التكرارات باستخدام قاعدة (أقل من) أو (أكبر من). ويعطي التكرار المتجمع طريقة سهلة لتجزئة التوزيع التكراري لمجموعة من الأجزاء المتساوية مثل المئينات أو العشيريات أو الربيعيات وأي متجزئات أخرى.

التكرار النسبي Relative frequency : عدد المرات التي يتكرر فيها حدث معين في عدد n محاولة في التجربة.

التوافيق Combinations : عدد الطرق المختلفة لترتيب المجموعة والتي لا تعتمد على ترتيب العناصر داخل المجموعة. وكل توفيق يحتوي على مجموعة مختلفة من العناصر.

التوزيع الاحتمالي Probability distribution : قائمة بالقيم المحتملة للمتغير العشوائي X والاحتمالات المصاحبة لها.

التوزيع التكراري Frequency distribution : تلخيص للبيانات باستخدام عدد من الفئات بحيث يكون لكل فئة طول محدد من الوحدات و يعبر عنها كفترة. ويعبر عمود التكرارات في الجدول عن عدد المشاهدات داخل كل فئة. ويجب أن يتساوى إجمالي التكرارات مع عدد المشاهدات.

توزيع المعاينة Sampling distribution : التوزيع الاحتمالي لمقدر أو إحصاءة مثل المتوسط. ويتم الحصول عليه بسحب جميع العينات الممكنة بحجم *n* من المجتمع ، ثم حساب المقدر (المتوسط) لكل عينة مسحوبة ، ثم تحديد الاحتمال المصاحب لكل قيمة من قيم المقدر.

التوزيع كبير المفرطح Platykurtic distribution : توزيع منبسط القمة مقارنة بالتوزيع مدبب الشكل.

ح

الحدث المركب Compound event : الحدث الذي يتكون من عدة أحداث متنافية الوقوع.

الحدث Event : الاسم الذي يعبر عن كل نتيجة للتجربة التي يمكن أن تقع في المحاولة الفردية. أيضاً ، أي مجموعة فرعية من المجموعة الشاملة للتجربة.

خ

الخطأ المعاينة Sampling error : الفرق بين قيمة الاحصاءة أو متغير العينة وقيمة معلمة المجتمع أو متغير المجتمع.

الخطاء المعياري للتقدير Standard error of the estimate : مقياس رقمي لتشتت النقاط حول معادلة الانحدار ؛ ويتم حسابه بأخذ الجذر التربيعي لمتوسط مربع الخطاء من تحليل التباين لخط الانحدار.

الخطاء المعياري للمتوسط Standard error of the mean : الانحراف المعياري لتوزيع المعاينة للمتوسط.

الخطاء من النـــوع الأول Type I error : الخطاء الذي يحدث برفض فرض العدم عندما يكون صحيح، ويكون باحتمال α.

الخطاء من النوع الثاني Type II error : الخطاء الذي يحدث بقبول فرض العدم عندما يكون خاطئ، ويكون باحتمال β.

د

دالة الكثافة الاحتمالية لتوزيع بواسون Poisson pdf : دالة كثافة احتمالية معرفة لمتوسط عدد مرات الحدوث لوحدة الزمن أو منطقة محددة وتعطي احتمال وقوع x في الفترة الزمنية القادمة أو المنطقة ؛ وهي دائماً ملتوية لليمين.

دالة الكثافة الاحتماليـــة لثنـــائي الحـــدين Binomial pdf : دالة الكثافة الاحتمالية المتحصل عليها من محاولات برنولي.

دالة الكثافة الاحتمالية للتوزيع الطبيعي Normal pdf : دالة كثافة احتمالية مستمرة للمتغير العشوائي x وتأخذ شكل الجرس بحدود تمتد من موجب ما لانهاية إلى سالب ما لانهاية ؛ وتتحدد بمتوسط المتغير وانحرافه المعياري.

دالة الكثافة الاحتمالية Probability density function (pdf) : دالة الكثافة الاحتمالية المنفصلة توضح الاحتمال لكل ناتج محتمل للمتغير العشوائي المنفصل. ودالة التوزيع الاحتمالي المتصلة دالة تعبر عن توزيع الاحتمالات لمدى من القيم للمتغير العشوائي المتصل.

درجات الحريـــة Degrees of freedom : عدد المشاهدات في العينة الحرة (المستقلة)

والقادرة على التغير وعليه يمكن اختيار أي قيمة منها.

س

السلسلة الزمنيـــة Time series : مجموعة من البيانات الإحصائية التي تم جمعها، وتسجيلها، أو ملاحظتها خلال فترات زمنية متوالية.

ش

شكل الانتشار Scatter diagram : رسم لنقاط قيم x و y ; يستخدم لتوضيح الارتباط بين المتغيرين، النمط في الخطاء العشوائي لنموذج الانحدار، ...إلخ.

ط

طريقة المربعـــات الـــصغرى Method of least squares : طريقة تركز على المتغير العشوائي y في تحليل البيانات وتعمل على تدنية مجموع مربعات الاختلافات في اتجاه y حول خط الانحدار. وتستخدم في تقدير معالم الانحدار a و b، والتي تمثل قيم القاطع والميل في معادلة الخط.

ع

العينات الاحتمالية Probability-based sample : أسلوب معاينة يعتمد على أن يكون

لكل عنصر في المجتمع فرصة معلومة بحيث يمكن شموله في العينة.

العينة الاجتهادية Judgment sampling : استخدام الاجتهاد الشخصي بدلاً من الطرق الموضوعية مثل الأرقام العشوائية لاختيار عناصر العينة من المجتمع.

العينة الحصصية Quota sampling : عينة يتم تصميمها لتكون مشابهة للمجتمع تبعاً لبعض الصفات الرئيسية. ويتم سحب العناصر من المجتمع حتى نصل لحصة تلك الصفة المختارة، ثم نختار عناصر تحتوي على صفات أخرى وهكذا.

العينة العشوائية Random sample : العينة التي يتم اختيارها بحيث يكون لكل مفردة في مجتمع المعاينة نفس الفرصة في الاختيار باحتمال متساوي لجميع المفردات.

العينة المنتظمة Systematic sample : يكون لدينا إمكانية الحصول على قائمة بجميع مفردات المجتمع والحصول على العينة باختيار كل kth وحدة في المجتمع، حيث k تشير إلى رقم صحيح يساوي تقريباً نسبة المعاينة $\frac{N}{n}$.

العينة Sample : مجموعة جزئية من المجتمع.

ف

الفئات الحقيقية Real classes : فئات في التوزيع التكراري يتم إنشائها بحيث يكون الحد الأعلى للفئة الأولى هو الحد الأدنى في الفئة الثانية، وهكذا. والفئات مستمرة لمدى البيانات ويجب أن تستخدم في الحسابات.

الفئات المنفصلة Stated classes : فئات في التوزيع التكراري يتم إنشائها بحيث تكون منفصلة بوحدة واحدة بين الحد الأعلى للفئة الأولى والحد الأدنى للفئة الثانية؛ مثلاً

٢١٥-٢٣٤ ، ٢٣٥-٢٥٤ ، ...إلخ ؛ وتستخدم في بعض الأحيان في العرض لأنها تؤدي لتقليل الالتباس في التفسير.

الفئات ذات النهايات المفتوحة Open-ended classes : فئات في التوزيع التكراري ليس لها نهاية محددة، مثل ٣٣٥ و أعلى، أكثر من ٣٥٠، ... إلخ ؛ وتستخدم عندما يكون هناك عدد قليل من الأرقام الكبيرة تنتشر خارج عرض الفئة مما يتطلب فئات كثيرة في حالة استخدام الفئات المعتادة بطول ثابت.

فترات الثقة Confidence interval : فترة تقدير لمعلمة المجتمع مبنية على الاحتمال، مثل μ ،ويتم إيجادها من توزيع المعاينة للمقدرة مثل $\bar{x}$.

فراغ العينة Sample space : مجموعة من نقاط العينة بحيث تمثل كل نقطة حدث واحد محتمل فقط ؛ وتشتمل المجموعة على جميع نقاط العينة.

فرض العدم Null hypothesis : عبارة إحصائية تعني عدم وجود فرق، أو لا يوجد تأثير، أو لا شي ؛ ويرمز له بالرمز *H*o.

الفروض الإحصائية Statistical hypothesis : الفروض التي يتم عملها حول بعض المعالم أو المتغيرات للمجتمع.

الفروض البديلة Alternative hypothesis:العبارة الإحصائية لقيم المجتمع والتي نعتقد بأنها صحيحة ونرمز لها بالرمز *Ha*أو *H1*.

ق

قاعدة سترجس Sturges's rule : صيغة رياضية تستخدم لحساب عدد الفئات k، وذلك عند تكوين التوزيع التكراري ؛ $k = 1 + 3.322(\log_{10} n)$ ، حيث n تمثل العدد

الكلي للتكرارات. وفي معظم الأحيان قد لاتكون قيمة k عدد صحيح وعليه يتم تقريبها لأعلى أو أدنى رقم صحيح.

القطاعات Blocks : طريق لحساب بعض مصادر الخطأ داخل البيانات وتخليصها من خطا التجربة وذلك بتقسيمها إلى قطاعات. ومثال على ذلك اختيار مجموعة من الحقول كقطاعات تختلف في محاولات التجربة نظراً لاحتوائها على أنواع مختلفة من التربة.

قوة الاختبار Power of the test : تعبر عن احتمال رفض فرض العدم عندما يكون خاطئ ؛ وتساوي ١ مطروحاً منه β ، احتمال الخطاء من النوع الثاني II، ويرغب الباحث أن تصل قوة الاختبار ١ لقيم الفروض البديلة التي محل اهتمامه.

القيمة الحرجة Critical value : قيمة جدوليه لاحصاءة الاختبار (z، t، ....)والتي تحدد قيمة α أو ($\alpha/2$) في طرف توزيع المعاينة لفرض العدم. فإذا كانت القيمة المحسوبة لاحصاءة الاختبار مساوية أو تزيد عن القيمة الحرجة فإننا نرفض فرض العدم.

القيمة المتوقعة أو التوقع الرياضي Expected value or mathematical expectation : المتوسط الحسابي للمتغير المرجح بالاحتمالات.، أي ترجيح كل قيم المتغير بالاحتمالات المصاحبة لها.

ك

الكفاءة Efficiency : يكون تصميم العينة أكثر كفاءة مقارنة بالتصاميم الأخرى إذا كانت نتائجه لها

الكفاية Sufficiency : يستخدم المقدر $\hat{\eta}$ جميع المعلومات المتاحة حول معلمة المجتمع

n المشمولة في بيانات العينة. وبالتالي فعند حساب متوسط العينة فإنه يتم استخدام جميع بيانات العينة.

م

المؤشر البسيط Simple index : مؤشر يعبر عن العلاقة بين رقمين بحيث يستخدم أحدهما كأساس.

المؤشر المركب Composite index : تجميع للمعلومات من عدة مجموعات من البيانات في موشر.

المتغير العشوائي Random variable : يشير إلى أي متغير تكون نتائجه مرتبطة بالصدفة أو العشوائية.

المتغير Variable : أي كائن يمكن أن يكون له مدى من القيم.

المتغيرات الكمية Quantitative variables : العناصر الممكن قياسها أو عدها.

المتغيرات المتصلة Continuous variables : نتيجة قياس ، وبالتالي فإنه يمكن افتراض أي قيمة في مدى المتغير بناء على دقة أداة القياس المستخدمة.

المتغيرات المنفصلة Discrete variables : المتغيرات التي تأخذ قيم رقمية صحيحة ، وهذه المتغيرات نتيجة للعد.

المتغيرات الوصفية Qualitative variables : المتغيرات التي تشير إلى بعض الصفات التي لا يمكن قياسها عددياً. والأمثلة تشمل الحالة الاجتماعية ، الأجزاء الصالحة والمعيبة ، أنواع الفواكه ، مجموعة العمر المنتمية إلى ، ... إلخ. هذا النوع من المتغيرات عادة يتم تعريفها بقيم افتراضية فمثلا نضع x=1 إذا كان لون الثور اسود وصفر فيما عدا ذلك.

متنافية الحدوث :**Mutually exclusive** وقوع حدث ما ينفي وقوع أي حدث آخر ؛ فمثلا ظهور الصورة عند رمي قطعة النقود مرة واحدة ينفي ظهور الكتابة في نفس الرمية وعليه فإن الصورة والكتابة أحداث متنافية.

المتوسط الحسابي Arithmetic Mean: مقياس النزعة المركزية الأكثر استعمالاً ، ويتم الحصول على متوسط المجتمع ، μ ، بجمع قيم المشاهدات ، X_i ، ثم قسمتها على عدده ، N .

متوسط المدى Midrange : متوسط يعبر عن المتوسط الحسابي لأقل قيمة و أكبر قيمة في مجموعة البيانات. وعند ترتيب البيانات فإنه عبارة عن متوسط أول وآخر قيمة. ويستخدم عادة في حالة البيانات التي يتم أخذ القراءات الكبرى والصغرى لها مثل متوسط درجات الحرارة اليومية و متوسط الأسعار اليومية للأسهم.

متوسط المربعات Mean square : مصطلح يعبر عن الفرق و يستخدم عادة في تحليل التباين

المتوسط المعدل Modified mean : يتم الحصول عليه باستعباد أكبر وأصغر قيمة في كل عمود ثم أخذ المتوسط لباقي القيم ؛ ويستخدم لحساب التغيرات الموسمية.

المتوسط Average: رقم يستخدم لعرض القيمة المتوسطة أو النزعة المركزية لمجموعة البيانات أو التوزيع.

المجتمع Population : جميع الوحدات في المجموعة موضع الدراسة.

مجموع المربعات Sum of squares : البسط في معادلة حساب التباين ؛ وهو يمثل مجموع مربعات الانحرافات للقيم عن المتوسط الحسابي.

المجموعة الشاملة Universal set : جميع الأحداث التي تشكل تجربة.

المجموعة Set : مجموعة من الأشياء معرفة وواضحة المعالم، وقد تكون حقيقية أو تخيلية.

محاولات برنولي Bernoulli trials : محاولات تحتوي على عمليات عشوائية لها نتيجتين محتملة فقط وليس لها نمط وقوع ثابت، واحتمال وقوع أي من تلك النتائج ثابت لكل محاولة.

المدرج التكراري Histogram : أعمدة بيانية للتوزيع التكراري؛ وترسم الأعمدة البيانية على المحور الأفقي x بحيث تكون حدود الفئات الحقيقية عند نهاية كل عمود، في حين تمثل التكرارات على المحور الراسي y. وإذا كانت الفئات غير متساوية الطول فلن يكون هناك أعمدة في المدرج التكراري.

المدى Range : مقياس للتشتت ويعرف على أنه الفرق بين بداية ونهاية التوزيع بعد ترتيب البيانات تصاعدياً أو تنازلياً.

المشاهدات Observations : مصطلح آخر للبيانات. ويشار إلى عناصر البيانات الفردية بالمشاهدات.

المصفوفة Array: مجموعة من القيم يتم ترتيبها تصاعديا من أصغر قيمة لأكبر قيمة.

المضلع التكراري Frequency polygon : منحنى خطي للتوزيع التكراري يوضح التكرارات على المحور الرأسي y ومراكز الفئات على المحور الأفقي x. ولتوصيل المنحنى بمحور x نرسم مركز لفئة تخيلية في كل نهاية للتوزيع لتكرار يساوي صفر. وبدون عمل ذلك فإن المنحنى يكون أعلى من المحور الأفقي ويكون شكله غير ملائم.

المعادلات الطبيعية Normal equations : المعادلات الناتجة من تطبيق طريقة المربعات الصغرى. بحيث يكون هناك معادلة طبيعية واحدة لكل معلمة انحدار ويمكن حل مجموعة المعادلات آنياً للحصول على القيم المقدرة للمعالم.

المعالم Parameters : متغيرات المجتمع.

معامل الاختلاف Coefficient of variation : المعامل الذي يحدد الاختلاف النسبي في مجموعة من البيانات.وهو عبارة عن نسبة الانحراف المعياري للمتوسط مضروبا في ١٠٠.

معامل الارتباط *r* *r*, Correlation coefficient : يقيس قوة درجة العلاقة بين المتغيرين العشوائيين. وتتراوح قيمته بين ± 1 أو $-1 < r < +1$

معامل تـــصحيح المجتمـــع المنتـــهى Finite population correction factor : معامل يستخدم للعينات الكبيرة نسبياً في المجتمعات الصغيرة عند حساب الخطأ المعياري للمتوسط ؛ وهو عبارة عن الجذر التربيعي لـ N- وهو عبارة عن حاصل قسمة n على N-1.

المعاملة Treatment : أي متغير عشوائي يتم التحكم به من قبل الباحث.

المعاينة الطبقية Stratified sampling : المعاينة التي يكون لدينا بعض المعلومات المسبقة عن المجتمع ويتم استخدام تلك المعلومات في تقسيم المجتمع إلى طبقات أو شرائح مختلفة. ويتم اختيار الشرائح بحيث تحتوي على خصائص متشابهه وتكون أكثر تجانس فيما بينها مقارنة بالمجتمع ككل. ويتم أخذ عينة من كل طبقة باستخدام طريقة العينة العشوائية البسيطة وتكون العينة الإجمالية للمجتمع عبارة عن العينات التي تم أخذها من الطبقات المختلفة.

المعاينة العنقودية Cluster Sampling : يتم أولاً اختيار مجموعات من الوحدات الفردية من المجتمع عشوائياً تسمى عناقيد، ثم يتم اختيار جميع العينات الفرعية للوحدات داخل كل عنقود للحصول على عناصر العينة الكلية.

المعاينة الملائمة Convenience sampling : عينة يتم اختيارها بناء على اختيارنا بغض النظر عن مدى تمثيلها للمجتمع.

المقدر المتسق Consistent estimator : تتركز $\hat{\eta}$ كليا عند قيمتها الفعلية كلما زاد حجم العينة ليصل ما لانهاية ، أي أن أصبح غير متحيز وتباينه يساوي الصفر. وكلما كان حجم العينة لا نهائي فإن المقدر المتسق $\hat{\eta}$ يعطي تقدير نقطة تام لـ η الفعلي.

المقدرات Estimators : المتغيرات العشوائية من العينة التي تم اختيارها لتقدير معالم المجتمع ، مثل استخدام $\bar{x}$ لتقدير μ .

المكملة Complement : تعني العدم ، وعليه فإن مكملة A ليست A ، ولكن جميع الحوادث الأخرى في المجموعة.

المنحنى الطبيعي القياسي Standard normal curve : توزيع طبيعي خاص له متوسط يساوي الصفر وانحراف معياري يساوي الواحد. احتمالات هذا التوزيع موجودة في جدول وبذلك يمكن تحويل جميع المنحنيات الطبيعية إلى هذا التوزيع باستخدام صيغة Z.

منفصلة Disjoint : مجموعتين ليس لها نفس العناصر. وتقاطعها يساوي صفر.

المنوال Mode : متوسط يحدد أكثر المشاهدات تكراراً في مجموعة البيانات. وفي حالة البيانات غير المبوبة ، فإننا ببساطة نختبر البيانات ونختار القيم من المشاهدات التي تتكرر غالباً.

ن

النتائج Outcomes : تعبر عن نتائج التجربة ، مثل عدد النقاط الممكنة التي تظهر للأعلى عند تجربة رمي زهرتي نرد معاً (٢ ، ٣ ، ٤ ، ... ، ١٢).

نفس الدرجة من الموثوقية وبتكاليف أقل. أيضا تشير الكفاءة إلى اختلافات المقدر.ويكون المقدر أكثر كفاءة مقارنة بالمقدرات الأخرى عندما يكون أقل تباين.

و

وحدة التجربة Experiment unit : الكيان الذي تتم عليه المعاملة

الوسط الحسابي المرجح Weighted arithmetic mean : يحسب باستخدام الأوزان لكل ملاحظة ماعدا الواحد. فمثلا ، يستخدم التكرار لترجيح كل مركز فئة عند حساب متوسط التوزيع التكراري.

الوسيط Median : متوسط مكان لا يتأثر بقيم المشاهدات في أطراف البيانات وعليه يعتبر أفضل متوسط يمكن استخدامه في حالة البيانات التي تكون فيها قيم متطرفة مثل بيانات الدخل والتعليم. وفي حالة البيانات غير المبوبة فإن المتوسط عبارة عن القيم التي في الوسط بعد ترتيب البيانات تصاعديا أو تنازلياً.

ثانياً: انجليزي – عربي

A

المؤشر Alternative hypothesis الفروض البديلـــة : العبـارة الإحـصائية لقـيم المجتمـع والتي نعتقد بأنها صحيحة ونرمز لها بالرمز Ha أو H1.

A priori Probabilities الاحتمالات القبلية: هي الاحتمـالات الـتي تحـدد باسـتخدام النظرية أو التقدير الحدسي وهي عكس التجريبي.

Arithmetic Mean المتوسط الحسابي:مقيـاس النزعـة المركزيـة الأكثر اسـتعمالاً ، ويـتم الحصول على متوسـط المجتمـع ، μ ، بجمـع قـيم المشاهدات ، X_i ، ثـم قـسمتها علـى عدده ، N .

Array المصفوفة : مجموعة من القيم يتم ترتيبها تصاعديا من أصغر قيمة لأكبر قيمة.

Average المتوسط :رقم يـستخدم لعـرض القيمـة المتوسـطة أو النزعـة المركزيـة لمجموعـة البيانات أو التوزيع.

B

Bernoulli trials محاولات برنولي :محاولات تحتوي على عمليات عشوائية لها نتيجـتين محتملة فقط وليس لها نمط وقوع ثابت ، واحتمال وقوع أي من تلك النتائج ثابـت لكـل محاولة.

Bais التحيز : يمكـن تطبيقـه علـى أي مقـدر ، وتكـون قيمتـه مـساوية للـصفر إذا كانـت القيمة المتوقعة للمقدر مساوية لمقدر المجتمع. مثلا ، إذا كان $E(\overline{x}) = \mu$ لمتوسـط العينـة ،

وما عدا ذلك فإن قيمة التحيز تساوي القيمة المتوقعة مطروحاً منها المقدرّة.

Binomial pdf دالة الكثافة الاحتمالية لثنائي الحدين : دالة الكثافة الاحتمالية المتحصل عليها من محاولات برنولي

Blocks القطاعات :طريق لحساب بعض مصادر الخطأ داخل البيانات وتخليصها من خطا التجربة وذلك بتقسيمها إلى قطاعات. ومثال على ذلك اختيار مجموعة من الحقول كقطاعات تختلف في محاولات التجربة ؛ نظراً لاحتوائها على أنواع مختلفة من التربة.

C

Cluster Sampling المعاينة العنقودية :يتم أولا اختيار مجموعات من الوحدات الفردية من المجتمع عشوائياً تسمى عناقيد ، ثم يتم اختيار جميع العينات الفرعية للوحدات داخل كل عنقود للحصول على عناصر العينة الكلية.

Coefficient of variation معامل الاختــلاف: المعامل الذي يحدد الاختلاف النسبي في مجموعة من البيانات.وهو عبارة عن نسبة الانحراف المعياري للمتوسط مضروبا في ١٠٠ .

Combinations التوافيق: عدد الطرق المختلفة لترتيب المجموعة والتي لا تعتمد على ترتيب العناصر داخل المجموعة. وكل توفيق يحتوي على مجموعة مختلفة من العناصر

Complement المكملة: تعني العدم ، وبالتالي فإن مكملة A ليست A ، ولكن جميع الحوادث الأخرى في المجموعة.

Composite index المؤشر المركب: تجميع للمعلومات من عدة مجموعات من البيانات في مؤشر

Compound event الحدث المركــب: الحدث الذي يتكون من عدة أحداث متنافية الوقوع.

Conditional probabilities الاحتمالات الشرطية :الاحتمالات المصاحبة للأحداث في المجتمعات الفرعية

Confidence interval فترات الثقة: فترة تقدير لمعلمة المجتمع مبنية على الاحتمال، مثل μ ، ويتم إيجادها من توزيع المعاينة للمقدرة مثل $\bar{x}$.

Consistent estimator المقدر المتسق:تتركز $\hat{\eta}$ كليا عند قيمتها الفعلية كلما زاد حجم العينة ليصل ما لانهاية ، أي أن أصبح غير متحيز وتباينه يساوي الصفر. وكلما كان حجم العينة لا نهائي فإن المقدر المتسق $\hat{\eta}$ يعطي تقدير نقطة تام لـ η الفعلي.

Continuous variables المتغيرات المتصلة: نتيجة قياس ، وبذلك فإنه يمكن افتراض أي قيمة في مدى المتغير بناء على دقة أداة القياس المستخدمة.

Convenience sampling المعاينة الملائمة: عينة يتم اختيارها بناء على اختيارنا بغض النظر عن مدى تمثيلها للمجتمع.

Correlation الارتباط:مقياس لدرجة العلاقة بين متغيرين عشوائيين ، x و y. وتكون مرتبطة بناء على حركتها معاً إما في اتجاه واحد أو اتجاه متعاكس.

Correlation coefficient , *r* معامل الارتباط *r*: يقيس قوة درجة العلاقة بين المتغيرين العشوائيين. وتتراوح قيمته بين ± 1 أو $-1 < r < +1$

Critical value القيمة الحرجة:قيمة جدوليه لاحصاءة الاختبار (z، t، ....)والتي تحدد قيمة α أو ($\alpha/2$) في طرف توزيع المعاينة لفرض العدم. فإذا كانت القيمة المحسوبة لاحصاءة الاختبار مساوية أو تزيد عن القيمة الحرجة فإننا نرفض فرض العدم.

Cross section البيانات القطاعية: مجموعة من البيانات الإحصائية التي تم جمعها خلال فترة زمنية محددة من مساحة واسعة.

Cyclical fluctuations التغيرات الدورية:حركة الدورات الاقتصادية التي تعكس

توسع الاقتصاد من فترة الكساد المنخفضة من خلال الازدهار الاقتصادي ثم بعد ذلك تنكمش باتجاه فترة كساد أخرى.

D

Degrees of freedom درجـــات الحريـــة:عدد المشاهدات في العينة الحرة (المستقلة) والقادرة على التغير وعليه يمكن اختيار أي قيمة منها.

Descriptive statistics الإحصاء الوصفي:بيانات رقمية.

Discrete variables المتغيرات المنفصلة:المتغيرات التي تأخذ قيم رقمية صحيحة ، وهذه المتغيرات نتيجة للعد.

Disjoint منفصلة :مجموعتين ليس لها نفس العناصر. وتقاطعها يساوي صفر.

Dispersion التشتت: مقياس لتمثيل أو صحة المتوسط ، ويعبر عن القيمة التي توضح مدى انتشار أو تركز القيم حول متوسطها.وكلما كان التشتت عن المتوسط صغير كلما كان المتوسط أكثر تمثيلاً وعليه نستطيع استخدام المتوسط كرقم لعرض كامل التوزيع.

F

Finite population correction factor معامل تصحيح المجتمع المنتهى :معامل يستخدم للعينات الكبيرة نسبياً في المجتمعات الصغيرة عند حساب الخطأ المعياري للمتوسط ؛ وهو عبارة عن الجذر التربيعي لـ N-   وهو عبارة عن حاصل قسمة n على N-1.

Frequency distribution التوزيع التكراري: تلخيص للبيانات باستخدام عدد من الفئات بحيث يكون لكل فئة طول محدد من الوحدات و يعبر عنها كفترة. ويعبر عمود

التكرارات في الجدول عن عدد المشاهدات داخل كل فئة. ويجب أن يتساوى إجمالي التكرارات مع عدد المشاهدات.

Frequency polygon المضلع التكـــراري: منحنى خطي للتوزيع التكراري يوضح التكرارات على المحور الرأسي y ومراكز الفئات على المحور الأفقي x. ولتوصيل المنحنى بمحور x نرسم مركز لفئة تخيلية في كل نهاية للتوزيع لتكرار يساوي صفر. وبدون عمل ذلك فإن المنحنى يكون أعلى من المحور الأفقي ويكون شكله غير ملائم.

H

Histogram المدرج التكـــراري :أعمدة بيانية للتوزيع التكراري ؛ وترسم الأعمدة البيانية على المحور الأفقي x بحيث تكون حدود الفئات الحقيقية عند نهاية كل عمود ، في حين تمثل التكرارات على المحور الراسي y. وإذا كانت الفئات غير متساوية الطول فلن يكون هناك أعمدة في المدرج التكراري.

I

Intersection التقاطع: المنطقة المشتركة لمجموعتين أو أكثر ، أو العناصر المشتركة لهذه المجموعات

J

Judgment sampling العينة الاجتهادية:استخدام الاجتهاد الشخصي بدلاً من الطرق

الموضوعية مثل الأرقام العشوائية لاختيار عناصر العينة من المجتمع.

M

Mean square متوسط المربعات: مصطلح يعبر عن الفرق و يستخدم عادة في تحليل التباين

Median الوسيط: متوسط مكان لا يتأثر بقيم المشاهدات في أطراف البيانات وعليه يعتبر أفضل متوسط يمكن استخدامه في حالة البيانات التي تكون فيها قيم متطرفة مثل بيانات الدخل والتعليم. وفي حالة البيانات غير المبوبة فإن المتوسط عبارة عن القيم التي في الوسط بعد ترتيب البيانات تصاعديا أو تنازلياً.

Method of least squares طريقة المربعات الصغرى: طريقة تركز على المتغير العشوائي y في تحليل البيانات وتعمل على تدنية مجموع مربعات الاختلافات في اتجاه y حول خط الانحدار. وتستخدم في تقدير معالم الانحدار a و b ، والتي تمثل قيم القاطع والميل في معادلة الخط.

Midrange متوسط المدى: متوسط يعبر عن المتوسط الحسابي لأقل قيمة و أكبر قيمة في مجموعة البيانات. وعند ترتيب البيانات فإنه عبارة عن متوسط أول وآخر قيمة. ويستخدم عادة في حالة البيانات التي يتم أخذ القراءات الكبرى والصغرى لها مثل متوسط درجات الحرارة اليومية و متوسط الأسعار اليومية للأسهم.

Mode المنوال: متوسط يحدد أكثر المشاهدات تكراراً في مجموعة البيانات. وفي حالة البيانات غير المبوبة ، فإننا ببساطة نختبر البيانات ونختار القيم من المشاهدات التي تتكرر غالباً.

Modified mean المتوسط المعدل :يتم الحصول عليه باستعباد أكبر وأصغر قيمة في كل عمود ثم أخذ المتوسط لباقي القيم ؛ ويستخدم لحساب التغيرات الموسمية.

Multicollinearity الارتباط الخطي :يحدث عندما تكون المتغيرات المستقلة في النموذج الخطي مرتبطة داخلياً بدرجة كبيرة.

Multiple linear regression الانحدار الخطي المتعدد :معادلة تشرح الاختلافات في y باستخدام مجموعة من المتغيرات المستقلة x (أكثر من واحد).

Mutually exclusive متنافية الحدوث :وقوع حدث ما ينفي وقوع أي حدث آخر ؛ فمثلا ظهور الصورة عند رمي قطعة النقود مرة واحدة ينفي ظهور الكتابة في نفس الرمية وعليه فإن الصورة والكتابة أحداث متنافية.

N

Nominal data بيانات اسمية: البيانات تكون أسمية فقط. فمثلا يتم تصنيف بيانات الأبقار حسب أسمائها ، عجول ، تبيعه ، تبيع ، أبقار ، ثيران... إلخ.

Nonparametric tests الاختبارات اللامعلمية :الاختبارات التي لا يكون هناك رغبة في وضع فروض حول شكل توزيع المعاينة للإحصاءات محل الاختبار.

Normal equations **المعادلات الطبيعية**: المعادلات الناتجة من تطبيق طريقة المربعات الصغرى. بحيث يكون هناك معادلة طبيعية واحدة لكل معلمة انحدار ويمكن حل مجموعة المعادلات آنياً للحصول على القيم المقدرة للمعالم.

***pdf* Normal *pdf* للتوزيع الطبيعي** :دالة كثافة احتمالية *pdf* مستمرة للمتغير العشوائي x وتأخذ شكل الجرس بحدود تمتد من موجب ما لانهاية إلى سالب ما لانهاية ؛ وتتحدد بمتوسط المتغير وانحرافه المعياري.

Null hypothesis فرض العدم: عبارة احصائية تعني عدم وجود فرق ، أو لا يوجد تأثير، أو لا شي ؛ ويرمز له بالرمز Ho.

O

Observations المشاهدات :مصطلح آخر للبيانات. ويشار إلى عناصر البيانات الفردية بالمشاهدات.

Ogive التكرار المتجمع: منحنى التكرار التجميعي ويكون على شكل حرف S حيث توضع التكرارات المتجمعة على المحور الرأسي y وتوضع الحدود العليا للفئات على المحور الأفقي x. ويمكن تجميع التكرارات باستخدام قاعدة (أقل من) أو (أكبر من). ويعطي التكرار المتجمع طريقة سهلة لتجزئة التوزيع التكراري لمجموعة من الأجزاء المتساوية مثل المئينات أو العشيريات أو الربيعيات وأي متجزئات أخرى.

Open-ended classes الفئات ذات النهايات المفتوحة :فئات في التوزيع التكراري ليس لها نهاية محددة، مثل ٣٣٥ و أعلى، أكثر من ٣٥٠... إلخ؛ وتستخدم عندما يكون هناك عدد قليل من الأرقام الكبيرة تنتشر خارج عرض الفئة مما يتطلب فئات كثيرة في حالة استخدام الفئات المعتادة بطول ثابت.

Ordinal data البيانات الترتيبية:بيانات تعبر عن علاقة الترتيب وبالتالي يمكن استخدام الرتب. فمثلاً المستهلك ربما يرتب تفضيله للطعام ، ... إلخ.

Outcomes النتائج:تعبر عن نتائج التجربة، مثل عدد النقاط الممكنة التي تظهر للأعلى عند تجربة رمي زهرتي نرد معاً (٢ ، ٣ ، ٤ ، ٠٠٠ ، ١٢).

P

Paired data البيانات المزدوجة: تكون البيانات غير المستقلة نظراً لوجود بعض

العوامل التي تجعل بينها علاقة ؛ فمثلاً البيانات المجموعة لمحاولات متنوعة لمزارع مختلفة عبارة عن بيانات غير مستقلة ؛ نظراً لأن المهارات الإدارية للمزارعين والظروف الحقلية...إلخ، مختلفة من مزرعة إلى مزرعة وبالتالي تؤثر على اختلاف الإنتاجية.

Parameters المعالم:متغيرات المجتمع.

Permutation التبديل :أي ترتيب تسلسلي لمجموعة أو مجموعة أشياء.

Platykurtic distribution التوزيع كبير المفرطح :توزيع منبسط القمة مقارنة بالتوزيع مدبب الشكل.

Point estimate : تقدير النقطة: قيمة واحدة للمقدر ؛ مثلاً ، $\bar{x} = 2$.

***Pdf* Poisson *pdf* لتوزيع بواســـون:** دالة كثافة احتمالية معرفة لمتوسط عدد مرات الحدوث لوحدة الزمن أو منطقة محددة وتعطي احتمال وقوع x في الفترة الزمنية القادمة أو المنطقة ؛ وهي دائماً ملتوية لليمين.

Population المجتمع: جميع الوحدات في المجموعة موضع الدراسة.

Power of the test قوة الاختبار: تعبر عن احتمال رفض فرض العدم عندما يكون خاطئ ؛ وتساوي ١ مطروحاً منه β ، احتمال الخطاء من النوع الثاني II، ويرغب الباحث أن تصل قوة الاختبار١ لقيم الفروض البديلة التي محل اهتمامه.

Probability الاحتمال:تعبر عن نسبة عدد الحوادث المفضلة لحدث A إلى عدد الحوادث في التجربة.

Probability-based sample العينات الاحتمالية:أسلوب معاينة يعتمد على أن يكون لكل عنصر في المجتمع فرصة معلومة بحيث يمكن شموله في العينة.

Probability density function (pdf) دالة الكثافة الاحتمالية:دالة الكثافة الاحتمالية المنفصلة توضح الاحتمال لكل ناتج محتمل للمتغير العشوائي المنفصل. دالة التوزيع

الاحتمالي المتصلة دالة تعبر عن توزيع الاحتمالات لمدى من القيم للمتغير العشوائي المتصل.

Probability distribution التوزيع الاحتمالي: قائمة بالقيم المحتملة للمتغير العشوائي X والاحتمالات المصاحبة لها.

Q

Qualitative variables المتغيرات الوصفية:المتغيرات التي تشير إلى بعض الصفات التي لا يمكن قياسها عددياً. والأمثلة تشمل الحالة الاجتماعية، الأجزاء الصالحة والمعيبة، أنواع الفواكه، مجموعة العمر المنتمية إلى ... إلخ. هذا النوع من المتغيرات عادة يتم تعريفها بقيم افتراضية فمثلا نضع x=1 إذا كان لون الثور اسود وصفر فيما عدا ذلك.

Quantitative variable المتغيرات الكمية:العناصر الممكن قياسها أو عدها.

Quartile deviation الانحراف الربيعي:مقياس للتشتت يستخدم فقط مع المتوسط؛ ويشير إلى تشتت البيانات في منتصف الوسط للتوزيع.

Quota sampling العينة الحصصية: عينة يتم تصميمها لتكون مشابهة للمجتمع تبعاً لبعض الصفات الرئيسية. ويتم سحب العناصر من المجتمع حتى نصل لحصة تلك الصفة المختارة، ثم نختار عناصر تحتوي على صفات أخرى وهكذا

R

Random sample **العينة العشوائية**:العينة التي يتم اختيارها بحيث يكون لكل مفردة في مجتمع المعاينة نفس الفرصة في الاختيار باحتمال متساوي لجميع المفردات.

Random variable المتغير العشوائي: يشير إلى أي متغير تكون نتائجه مرتبطة بالصدفة أو العشوائية.

Range المدى: مقياس للتشتت ويعرف على أنه الفرق بين بداية ونهاية التوزيع بعد ترتيب البيانات تصاعدياً أو تنازلياً.

Rank correlation ارتباط الرتب: يحسب الارتباط العادي باستخدام الرتب للبيانات بدلاً من المشاهدات الأصلية.

Real classes الفئات الحقيقية : فئات في التوزيع التكراري يتم إنشائها بحيث يكون الحد الأعلى للفئة الأولى هو الحد الأدنى في الفئة الثانية ، وهكذا. والفئات مستمرة لمدى البيانات ويجب أن تستخدم في الحسابات.

Relative frequency التكرار النسبي: عدد المرات التي يتكرر فيها حدث معين في عدد n محاولة في التجربة.

Residual البواقي : الفرق بين القيمة المشاهدة لـ Y والقيمة المقدرة لـ Y من خط الانحدار ، ويرمز لها بالرمز e_i.

S

Sample العينة: مجموعة جزئية من المجتمع.

Sample space : فراغ العينة: مجموعة من نقاط العينة بحيث تمثل كل نقطة حدث واحد محتمل فقط ؛ وتشتمل المجموعة على جميع نقاط العينة.

Sampling distribution توزيع المعاينة: التوزيع الاحتمالي لمقدر أو إحصاءة مثل المتوسط. ويتم الحصول عليه بسحب جميع العينات الممكنة بحجم n من المجتمع ، ثم حساب المقدر (المتوسط) لكل عينة مسحوبة ، ثم تحديد الاحتمال المصاحب لكل قيمة من قيم المقدر.

Sampling error خطأ المعاينة : الفرق بين قيمة الاحصاءة أو متغير العينة وقيمة معلمة المجتمع أو متغير المجتمع.

Scatter diagram شكل الانتشار: رسم لنقاط قيم x و y ; يستخدم لتوضيح الارتباط بين المتغيرين، النمط في الخطاء العشوائي لنموذج الانحدار، ... إلخ.

Seasonal variation الـــتغيرات الموسميـــة: دورات قصيرة الأجل تحدث للبيانات وتستكمل دورتها في غضون السنة ، وتتكرر سنويا.

Secular trend الاتجاه الزمني: حركة انسيابية تصاعدية أو تنازلية تميز السلسلة الزمنية خلال فترة طويلة من الزمن، مثل ١٠ سنوات أو أكثر.

Set المجموعة: مجموعة من الأشياء معرفة وواضحة المعالم، وقد تكون حقيقية أو تخيلية.

Simple index المؤشر البـــسيط: مؤشر يعبر عن العلاقة بين رقمين بحيث يستخدم أحدهما كأساس.

Skewness الالتواء : التماثل النسبي للتوزيع. إذا كان التوزيع متماثل الشكل تماماً فليس هناك التواء. إما إذا كان للتوزيع طرف أيمن فإنه ذو التواء موجب أو ملتو لليمين، وإذا كان له طرف في اليسار فإنه ذو التواء سالب أو ملتو لليسار.

Standard deviation الانحـــراف المعيـــاري: مقياس للتشتت يستخدم مع المتوسط الحسابي. ويتم حساب قيمته بناء على جميع المشاهدات في مجموعة البيانات. ويحسب بأخذ الجذر التربيعي للمتوسط الحسابي لمربعات انحراف القيم عن متوسطها.

Standard error of the estimate الخطاء المعياري للتقـــدير : مقياس رقمي لتشتت النقاط حول معادلة الانحدار؛ ويتم حسابه بأخذ الجذر التربيعي لمتوسط مربع الخطاء من تحليل التباين لخط الانحدار.

Standard error of the mean الخطاء المعياري للمتوسط: الانحراف المعياري لتوزيع المعاينة للمتوسط.

Standard normal curve المنحنى الطبيعي القياسي: توزيع طبيعي خاص له متوسط يساوي الصفر وانحراف معياري يساوي الواحد. احتمالات هذا التوزيع موجودة في الجدول وبالتالي يمكن تحويل جميع المنحنيات الطبيعية إلى هذا التوزيع باستخدام صيغة Z.

Stated classes الفئات المنفصلة: فئات في التوزيع التكراري يتم إنشائها بحيث تكون منفصلة بوحدة واحدة بين الحد الأعلى للفئة الأولى والحد الأدنى للفئة الثانية ؛ مثلاً ٢١٥-٢٣٤ ، ٢٣٥-٢٥٤ ، ... إلخ ؛ وتستخدم في بعض الأحيان في العرض لأنها تؤدي لتقليل الالتباس في التفسير.

Statistical hypothesis الفروض الإحصائية: الفروض التي يتم عملها حول بعض المعالم أو المتغيرات للمجتمع.

Statistical inference الاستدلال الإحصائي: تطوير وتطبيق الأساليب والتقنيات المتعلقة بجمع وتلخيص وتفسير البيانات العددية ؛ لذا فإن الاعتماد على الخلاصة المرسومة من هذه الإجراءات يمكن تقييمه بصورة موضوعية باستخدام العبارات الاحتمالية.

Statistics الإحصاءات : متغيرات العينة.

Stratified sampling المعاينة الطبقية: المعاينة التي يكون لدينا بعض المعلومات المسبقة عن المجتمع ويتم استخدام تلك المعلومات في تقسيم المجتمع إلى طبقات أو شرائح مختلفة. ويتم اختيار الشرائح بحيث تحتوي على خصائص متشابهه وتكون أكثر تجانس فيما بينها مقارنة بالمجتمع ككل. ويتم أخذ عينة من كل طبقة باستخدام طريقة العينة العشوائية البسيطة وتكون العينة الإجمالية للمجتمع عبارة عن العينات التي تم أخذها من الطبقات المختلفة.

Sturges's rule قاعدة ستـرجس: صيغة رياضية تستخدم لحساب عدد الفئات k، وذلك عند تكوين التوزيع التكراري ؛ $k = 1 + 3.322(\log_{10} n)$ ، حيث n تمثل العدد الكلي للتكرارات. وفي معظم الأحيان قد لاتكون قيمة k عدد صحيح وعليه يتم تقريبها لأعلى أو أدنى رقم صحيح.

Sufficiency الكفاية: يستخدم المقدر $\hat{\eta}$ جميع المعلومات المتاحة حول معلمة المجتمع η المشمولة في بيانات العينة. وبذلك فعند حساب متوسط العينة فإنه يتم استخدام جميع بيانات العينة.

Sum of squares مجموع المربعات: البسط في معادلة حساب التباين ؛ وهو يمثل مجموع مربعات الانحرافات للقيم عن المتوسط الحسابي.

Systematic sample العينة المنتظمة: يكون لدينا إمكانية الحصول على قائمة بجميع مفردات المجتمع والحصول على العينة باختيار كل kth وحدة في المجتمع، حيث k تشير إلى رقم صحيح يساوي تقريباً نسبة المعاينة $\frac{N}{n}$.

T

Time series السلسلة الزمنية: مجموعة من البيانات الإحصائية التي تم جمعها، وتسجيلها، أو ملاحظتها خلال فترات زمنية متوالية.

Treatment المعاملة: أي متغير عشوائي يتم التحكم به من قبل الباحث.

Type I error الخطاء من النوع الأول: الخطاء الذي يحدث برفض فرض العدم عندما يكون صحيح، ويكون باحتمال α.

Type II error الخطاء من النوع الثاني: الخطاء الذي يحدث بقبول فرض العدم عندما يكون خاطئ، ويكون باحتمال β .

U

Union الاتحاد: العناصر الواردة في مجموعة واحدة أو مجموعة ثانية أو تلك المشتركة في كلتا المجموعتين.

Universal set المجموعة الشاملة: جميع الأحداث التي تشكل تجربة.

V

Variable المتغير:أي كائن يمكن أن يكون له مدى من القيم.

Variance التباين: مربع الانحراف المعياري، وأيضاً يسمى متوسط المربعات في تحليل التباين. ويجب أن نستخدم التباين في الحسابات الرياضية، ثم بعد ذلك نحسب الانحراف المعياري إذا كان هو المطلوب.

Variation within samples الاختلاف داخل العينة:مصطلح يستخدم عادة في تحليل التباين باتجاه واحد للإشارة إلى مجموع مربعات الخطاء أو متوسط مربعات الخطاء.

W

Weighted arithmetic mean الوسط الحسابي المرجح: يحسب باستخدام الأوزان لكل ملاحظة ماعدا الواحد. فمثلا، يستخدم التكرار لترجيح كل مركز فئة عند حساب متوسط التوزيع التكراري.

كشاف الموضوعات

ب

ت

ج

ن

و

نبذة عن المؤلف

الأستاذ الدكتور/ مهدي بن معيض السلطـــان

أستاذ الاقتصاد القياسي والإحصاء التطبيقي

قسم الاقتصاد الزراعي–كلية علوم الأغذية والزراعة

جامعة الملك سعود – الرياض

ALSULTAN@KSU.EDU.SA

- من مواليد عام ١٣٨٦هـ بقرية الغرس بمنطقة عسير. التحق بجامعة الملك سعود ، كلية علوم الأغذية والزراعة ، قسم الهندسة الزراعية عام ١٤٠٥هـ (١٩٨٥م) وتخرج فيها ١٤٠٨عام هـ بتقدير جيد جداً.
- عمل بوزارة الزراعة باحثاً اقتصاديا للفترة من ١٤٠٨- ١٤١٦هـ.
- التحق ببرنامج الماجستير بقسم الاقتصاد الزراعي عام ١٤١٤هـ ثم عين معيداً بالقسم عام ١٤١٦هـ.
- حصل على الماجستير في الاقتصاد الزراعي من قسم الاقتصاد الزراعي بجامعة الملك سعود بتقدير ممتاز عام ١٤١٨هـ ثم عين محاضر بالقسم.
- ابتعث إلى الولايات المتحدة الأمريكية عام ١٤١٩هـ للحصول على الدكتوراه، وحصل على درجة الدكتوراه في الاقتصاد القياسي من قسم الاقتصاد والموارد الطبيعية بتقدير ممتاز، جامعة ولاية كلورادو الحكومية بمدينة فورت كولنز، ولاية كلورادو ، الولايات المتحدة الأمريكية عام ١٤٢٣هـ(٢٠٠٢م).
- عين أستاذ مساعد بقسم الاقتصاد الزراعي، كلية علوم الأغذية والزراعة عام ١٤٢٣ هـ ثم رقي لرتبة استاذ مشارك عام ١٤٢٧هـ ثم إلى رتبة أستاذ عام ١٤٣١هـ.
- اهتماماته البحثية في مجال الاقتصاد التطبيقي، الاقتصاد القياسي واقتصاديات البيئة والموارد الطبيعية وتطبيقات الإحصاء في مجال الاقتصاد والبيئة والزراعة.
- شارك في عضوية كثير من اللجان على مستوى القسم والكلية والجامعة ولديه إسهامات متعددة في خدمة المجتمع.
- عضو جمعية الاقتصاديين الزراعيين الأمريكية وعضو بجمعية الاقتصاد السعودية والجمعية السعودية للعلوم الزراعية.
- نشر العديد من الأبحاث في المجلات المحلية والإقليمية والدولية وشارك في العديد من المشاريع البحثية الوطنية.
- أشرف على العديد من رسائل الماجستير وممتحن لرسائل أخرى.
- شارك في تقديم العديد من الدورات والبرامج التدريبية باستخدام البرامج الإحصائية الحديثة.

www.ksu.edu.sa

ردمك: ٦-٩٠٥-٥٥-٩٩٦٠-٩٧٨
ISBN: 987-9960-55-905-6